Herrebrøden, H., Espeseth, T., & Bishop, L. (2023). Mental Effort in Elite and Nonelite Rowers. Journal of Sport and Exercise Psychology, 45(4), 208-223. https://doi.org/10.1123/jsep.2022-0164

Publisher: Human Kinetics

Mental effort in elite and non-elite rowers

Henrik Herrebrøden[a,b*], Thomas Espeseth[b,c], & Laura Bishop[a,d]

[a]*RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Oslo, Norway*

[b]*Department of Psychology, University of Oslo, Oslo, Norway*

[c]*Department of Psychology, Oslo New University College, Oslo, Norway*

[d]*Department of Musicology, University of Oslo, Oslo, Norway*

*Correspondence: Henrik Herrebrøden, henrikh@henrikh.no

Abstract

Mental effort (intensity of attention) in elite sports has remained a debated topic and a challenging phenomenon to measure. Thus, a quasi-ecological laboratory study was conducted to investigate mental effort in elite rowers as compared to a group of non-elites. Findings suggest that eye-tracking measures–specifically, blink rates and pupil size–can serve as valid indicators of mental effort in physically demanding sports tasks. Further, findings contradict the notion that elite athletes spend less cognitive effort than their lower-level peers. Specifically, elites displayed similar levels of self-reported effort and performance decrement with increasing mental load, and significantly more mental effort overall as measured by pupil size increase (relative to baseline) during rowing trials, as compared to the non-elites in the sample. Future studies on eye tracking in sports may include investigations of mental effort in addition to selective attention during physically demanding tasks.

Keywords: attention; expertise; rowing; pupillometry; blinks; dual task.

Introduction

Cognitive psychology has a strong tradition for studying *selective* aspects of attention (e.g., which opponent the footballer attends to), but the last decades have seen increasing interest in *intensive* aspects of attention, or mental effort (e.g., how intensely the footballer attends to an opponent; Kahneman, 1973). Intriguingly, high intensity has not received the same favorable treatment in the cognitive domain as it has in the physical. On the contrary, automaticity and the ability to perform with less cognitive effort is regarded as a hallmark of expertise (Fitts & Posner, 1967). Certain views of elite athletes portray them as *zombies* (Breivik, 2013), who move around with little or no conscious thinking, in a sleep-walk type state. Indeed, skilled athletes are frequently advised to reduce thinking or deliberate attention, and even to "play 'outside your head' or at least your prefrontal cortex" (Beilock, 2011, p. 198). Too much attention to the task at hand, they are warned, may lead to *choking* or *paralysis by analysis* (Beilock, 2011).

The low-effort recommendations find support from three main sources. First, self-report studies, such as qualitative research on *flow* (Csikszentmihalyi, 2000), suggest that athletes often experience calmness and little conscious thinking during excellent performances (e.g., Chavez, 2008; Jackson, 1996). Second, dual-task performance studies suggest that skilled athletes may perform well while their minds are occupied with a secondary cognitive task (Beilock et al., 2002; Gray, 2004). Third, psychophysiological studies suggest that practice over time enables *neural efficiency* (Del Percio et al., 2008) and *hypofrontality* (Dietrich, 2004), namely the ability to perform with less cortical activation in general and less frontal cortex activation in particular (for a review of sport studies, see Filho et al., 2021).

However, there are numerous shortcomings in past research on this topic. First, research on elite athletes, especially at the highest levels (Swann et al., 2015), is scarce. Second, a limited number of sports tasks have been investigated, as studies are often conducted on golf putting and other tasks with relatively low demands in terms of physical effort (Perrey & Besson, 2018). Third, the tasks involved–especially in lab contexts–often have questionable ecological validity (for a review, see Christensen et al., 2015). These limitations reflect the fact that capturing cognition during sports performance is a methodological challenge. However, recent years have seen interesting developments in this regard. For instance, Whitehead et al. (2015, 2019) have conducted think-aloud research where athletes verbalize what they think in real-time while executing their skills. Findings from such studies suggest that cognition is nuanced, as performers report different kinds of cognition during different stages of their performance. Recent theoretical developments also reflect more nuanced views, as compared to the low-effort zombie view of skilled performance. The *Mesh* model (Christensen et al., 2016), for example, suggests an interplay between low-effort and high-effort cognitive processes. Specifically, experts may depend more on cognitive control–and less on automaticity–as task conditions become more difficult. Similarly, models have suggested that experts may perform well with different approaches, both when mental effort is relatively low and also when it is relatively high (Bertollo et al., 2016; Swann et al., 2016, 2017). Hence, continuous measurement of cognition during the actual performance, for example via think-aloud, is needed to get an accurate view of cognitive effort across different task conditions.

Yet, think-aloud and similar verbal methods may have their shortcomings. First, it may not be practical to verbally report on one's thought process during every kind of sporting event, especially during physically strenuous tasks. Second, athletes may experience cognition that is non-reportable (Schooler et al., 1993) or at least hard to put into words. Herein lies the appeal of neuroscientific or psychophysiological methods such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), as we may gain insight into the performers' cognition-related physiology while they do not have to report anything. In recent years, a non-invasive alternative has been increasingly used in sports research, namely eye tracking.

Eye behavior is often used to study selective attention (e.g., where does the footballer look?), but it can also reveal something about the mental effort involved in task performance. For example, how often we blink can indicate our cognitive load. Blink rates have tended to go down during visually demanding tasks (e.g., driving on a curvy road) and up during more internal processing (e.g., calculating) (Marquart et al., 2015). While research is scarce, the sports community has taken note of blinks as a potentially relevant factor. Aksel Lund Svindal, one of the best alpine skiers of all time, is notorious for not blinking during a two-minute downhill race (Red Bull, 2017). Studies have indeed found lower blink rates in elite athletes compared to non-sporting controls during computerized response tasks with visual stimuli, with elite samples from swimming (Pei et al., 2021) and women's cricket (Barrett et al., 2020). However, the latter study found no difference in blink rates between male elite athletes and male controls (Barrett et al., 2020). The role of blinking during actual sports performance remains relatively unexplored and unsettled.

The most direct measure of mental effort may be *pupillometry*, or pupil size measurement (Kahneman, 1973). Our pupils are affected not only by light conditions but also by mental states and task demands. A change in pupil size, typically a dilation (increased diameter), is a robust indicator of increased mental load in tasks involving math problems (Hess & Polt, 1964), language processing (Just & Carpenter, 1993), multiple object tracking (Alnæs et al., 2014), and more. Pupil size indexes activity in the locus coeruleus in the brainstem, which is connected to a broad range of cortical regions and involved in numerous cognitive operations such as controlling attention (Alnæs et al., 2014; Laeng et al., 2012). Pupillometry can be used as a global measure of cortical activity and the arousal associated with cognitive effort (Just et al., 2003; Kahneman, 1973; Larsen & Waters, 2018), and is therefore suited to test the zombie hypothesis in sport. If this hypothesis is true and expert athletes do not rely on mental effort during their performance, one should observe little pupil size change during task execution compared to resting states. In recent years, sports studies have employed pupillometry and contradicted this prediction.

Campbell et al. (2019) found significant increases in pupil size during a putting task, as compared to baseline, in golfers. Further, peak pupil sizes occurred at the onset of Quiet Eye (QE; Vickers, 2016), namely a period where the performer fixates on a target (in this case, the golf ball) before executing a key movement. These results were interpreted as evidence of the fact that golf putting, and especially QE, is cognitively demanding. The observation that QE promotes changes in pupil diameter has been replicated in other recent studies employing golf putting (Carnegie et al., 2020) and darts (Simpson et al., 2022). However, the latter studies found pupil *constrictions* (reduced diameter) to be indicative of mental effort during QE, instead of pupil *dilations*. In any event, a pupil that either shrinks or grows in size, depending on task characteristics (see Fletcher et al., 2017), appears to be a promising indicator of mental effort in sports.

Psychophysiological sports studies have frequently used tasks with limited physical demands, such as golf putting or darts. This is clearly practical since these activities will naturally invite participants to keep their heads still while brain activity is measured. Yet, if methods such as pupillometry should mark their place as universal tools for motor tasks, and if findings are to be generalizable to the broader sports community, measurements during physically strenuous tasks should be the next step. This seems particularly important since physical effort has its own effect on pupil size. For example, Hayashi et al. (2010) observed increased pupil sizes as participants invested more physical effort on an ergometer bike. The largest pupil sizes in this study were measured during the highest intensity level, with more than 140 heartbeats per minute, which is a typical heart rate level for endurance training at moderate intensities. Zénon et al. (2014) found a similar effect during a grip task, with participants' pupil size corresponding to both objective and subjective measures of physical effort. A study that varies mental and physical load during motor performance has yet to be conducted and will be important to establish the respective impact of different kinds of effort on pupil size (Bishop et al., 2021).

The present study

The current study aimed to investigate mental effort in athletes at the highest elite levels and, simultaneously, test the feasibility of using eye-tracking data in a physically strenuous sports task. Thus, rowers were invited to use a dynamic rowing ergometer–designed to mimic on-water rowing mechanics–under varying conditions in terms of mental and physical load. In a within-subjects design, participants were asked to row at a constant pace (low mental load), row at a constant pace while using a "race plan", namely focusing on self-selected and rowing-related task cues (medium mental load), or row at a constant pace while solving math problems (high mental load). These conditions were conducted twice, rowing at 75% of max physical intensity (low physical load) and 85% of max (high physical load). Male members from the Olympic group of the Norwegian national rowing team, preparing for the 2020 games in Tokyo, were recruited as an elite participant group. We also recruited a non-elite group, consisting of rowers with lower levels of experience and performance records, for comparison.

Mental effort was measured in several different ways. Physiological measures were obtained via eye tracking, providing blink rates and pupil size data. Performance indicators of mental effort were also included, as the dual-task conditions allowed us to see how attending to an external task would affect the rowing component. Finally, self-reports were included, as rowers rated effort-related aspects of their performance via a questionnaire. In addition to testing the zombie hypothesis by investigating mental effort in elites and non-elites, the current experimental approach provided an exploration of the relationship between mental and physical aspects of effort in motor performance, and also the link between objective and subjective measures of mental effort.

Method

Participants

The elite group consisted of nine male rowers, all part of the highest-level training group in the Norwegian national rowing team (Team Norway). Most of the rowers were qualified and preparing for the 2020 Olympic Games in Tokyo. The remaining elites were high-level rowers that practiced with Team Norway, despite not being qualified for the Olympic games at the time of recruitment. The recruited elite rowers could be categorized as Competitive Elite, Successful Elite, and World-class Elite, respectively, based on Swann et al's (2015) classification system and their career achievements (as listed on www.worldrowing.com). A Team Norway head coach estimated that this training group spent 900 hours on physical training (approx. 17 hours per week, on average) in 2020.

The non-elites were nine male rowers who had varying degrees of experience. Criteria for being in this group was that they had never represented Norway’s official national team as a rower, nor won an individual medal in any major national or international competition. Further, they needed to provide a result from a 2-kilometer (2K) rowing ergometer race or test to be included in the present study. This latter criterion was included because we wanted to control participants' physical effort by asking them to row at certain speeds based on their respective maximal capacities (see later description of physical load). When asked how much they had trained physically per week, on average, over the last 365 days, all non-elites estimated a minimum of three sessions or a minimum of 7 hours of weekly physical training average.

All rowers voluntarily agreed to participate at a proposed time that fit with their training schedule. No reward or reimbursement was offered to them. Ahead of participation, a brief interview (approx. 15 minutes) was conducted with each rower to exchange relevant information, and informed consent forms were signed. The Norwegian Centre for Research Data (NSD) approved this study (project identification number 455008).

Materials and measurements

The performance context. The study was conducted in a motion capture laboratory at the University of Oslo, Norway. The measured light conditions were 268.5 LUX and kept the same for all participants. A Row Perfect 3 (RP3) Model S was used for rowing. This is a *dynamic* ergometer, providing a closer resemblance to on-water rowing than *static* ergometers (Kleshnev, 2005).

Eye tracking. Two different mobile, head-mounted eye trackers were used. The first eight participants, all from the elite group, were tested with SensoMotoric Instruments (SMI) Eye Tracking Glasses 2 Wireless and iViewETG software, version 2.7.1 (SensoMotoric Instruments Inc, Teltow, Germany), recording each eye with infrared cameras at 60 Hz. The glasses were connected to a Samsung phone with software installed, and we made sure that the cord was not interfering with the rower's movements by attaching it to a string hanging from the ceiling, with the phone stored at a safe distance behind the rower. A three-point calibration was conducted ahead of each participant's warmup and first trial. The rowers' eye movements were recorded continuously throughout all trials unless the rower needed to leave the lab for a break.

Due to technical difficulties[1], the eye trackers had to be switched from the ninth participant and onwards. This change, and the comparability of the two eye trackers, are further discussed in Appendix A. Pupil Labs' Pupil Core glasses with Pupil Capture software, version 3.4.0 (Pupil Labs GmbH, Berlin, Germany) was used for the remaining participants. These Pupil Labs glasses resemble the SMI glasses by using infrared cameras and providing pupil size information in millimeters, albeit with a higher recording frequency of 120 Hz and with adjustable camera positioning (whereas SMI uses fixed camera positions built into the glasses). The setup was similar to the one used for previous participants, with glasses connected to a laptop behind the rower via a chord. A five-point calibration was conducted ahead of each participant's warmup and before every trial.

Self-report. To get subjective ratings of effort, the National Aeronautics and Space Administration-Task Load Index (NASA-TLX; Hart & Staveland, 1988) was used. NASA-TLX is a multi-dimensional scale allowing participants to give visual analogue scale ratings of the following items regarding any given task: Mental demand, Physical demand, Temporal demand (i.e., time pressure), Performance, Effort, Frustration.

Other measurements. We used various equipment to collect data that will not be the primary focus of this current article. The same Polar heart rate belt and watch were used by the participants, with the exception of two elites who preferred to use their own Polar equipment. Motion data were captured with a Qualisys motion capture system, with markers placed on various parts of the rower's body and the ergometer. Electromyography (EMG) data were collected via a Delsys system with sensors placed on various muscle groups. A Canon XF105 video camera recorded the rowers' movements from the side during trials. Finally, sound was either recorded through the microphones in the SMI glasses or via a Røde Wireless Go microphone when using the Pupil Labs glasses.

Task

Each rowing trial lasted three minutes. There were six different conditions that varied in terms of mental and physical load.

Mental load (ML). Three different levels of ML were used. With low ML, the rowers were simply asked to maintain a constant, agreed-upon split (i.e., time spent per 500 meters, given the current pace) throughout the trial. With medium ML, the rowers were asked to use a self-composed race plan (for example by thinking of various technical cues) while maintaining a constant pace. Specifically, they were given the following instructions: "Row for three minutes with the agreed-upon split. Additionally, we want you to use a race plan, while keeping a constant split. Your race plan may for example contain one or more technical cues, and/or counting, that you could have used in a rowing ergometer race. The goal of this condition is to simulate race mentality, while keeping a constant split." Then participants were asked to state their race plan orally before starting their trial. Although the low ML and medium ML conditions had similar absolute requirements (i.e., to maintain a constant split), participants' oral statements indicated that they involved different attentional strategies[2], and subjective reports suggested that the medium ML conditions were more mentally demanding than the low ML conditions (see the Results section).

However, it is not clear, based on past research, how mentally demanding it is to focus on task-related cues in sports and how this compares to other types of (cognitive) task performance. To explore this further and aid interpretation of findings, a manipulation that involved cognitive problem-solving during rowing performance was added. Hence, in the high ML condition, participants performed a dual task by responding to arithmetic (math) problems while rowing at a constant pace. The math problems were tasks involving addition (based on Zarjam et al., 2012) and multiplication (based on Ahern & Beatty, 1979) of varying difficulty. Specifically, each participant was asked to solve the following types of problems, in said order, while rowing:

1. Addition of one- and two-digit numbers (e.g., 35 + 2)
2. Addition of one- and two-digit numbers with one carry (e.g., 63 + 9)
3. Addition of two-digit numbers with one carry (e.g., 73 + 42)
4. Multiplication of [digits 6, 7, 8, or 9] by [digits, 12, 13, or 14] (e.g., 8 x 12)
5. Multiplication of [digits 6, 7, 8, or 9] by [digits 16, 17, 18, or 19] (e.g., 9 x 16)
6. Multiplication of [digits 11, 12, 13, or 14] by [digits 16, 17, 18 or 19] (e.g., 15 x 16)

A rowing trial in each dual-task condition consisted of two rounds of such math problems of increasing difficulty, so that a total of 12 problems were solved per trial. The same rower was never given the same problem twice, and the same problems were used for each participant. Each problem was presented via speakers, from pre-recorded sound files. Participants had approximately 10 s to respond before the next problem was presented. A short beep sounded before they were read a new problem.

Physical load (PL). In addition to ML manipulations, participants were instructed to maintain a steady pace with two different physical intensity levels: low and high. Specific pace levels were proposed for each individual rower based on an estimate of expected 2K race time, given their personal best times and current form. With low PL, rowers were told to row with approximately 75% of their expected 2K split time. With high PL, they adhered to approximately 85% of their expected 2K split time. Participants followed the agreed-upon split times by monitoring a screen in front of them with stroke-by-stroke information while rowing. The reason for keeping the pace constant during trials was to control the physical effort across the different levels of ML, since both physical and mental effort affects pupil size. Participants in the current study did not use the higher intensity zones (approaching max intensity) as this might have caused considerable noise to the physiological recordings.

In total, the experiment had a 3 x 2 task design, summarized in Table 1. All rowers participated in all six conditions[3]. Each rower alternated between conditions of low and high physical intensity, to avoid fatigue. Partial counterbalancing was used to ensure that no rower within each group had the same order of conditions. The condition order was matched between groups.

Table 1

The Design and Loads (i.e., Manipulations) Used in the Current Study.

	Low ML	**Medium ML**	**High ML**
Low PL	Rowing (75% of max)	Rowing (75% of max) with a race plan	Rowing (75% of max) and math
High PL	Rowing (85% of max)	Rowing (85% of max) with a race plan	Rowing (85% of max) and math

Procedure

After arriving in the laboratory, participants first provided some basic information via questionnaires. They were presented with proposed splits (pace values) that they could follow in the different intensity zones. If these numbers seemed appropriate, the rowers chose the option of accepting them. They were also allowed to adjust the numbers, if they felt the suggested pace was inappropriate (due to their current form, fatigue level, or other factors). Most rowers accepted the proposed split times, while a few made minor tweaks to the suggestions.

Once training gear was put on and all the equipment was in place, the rowers were allowed to warm up freely according to their own warm-up routine. A sound check was conducted while rowing at the beginning of the warmup, by rowers reading back a string of three numbers presented via sound speakers, to make sure they would be able to hear the math problems that were to be presented in the dual-task conditions. After the warmup, participants received general instructions, including encouragement to row while looking at the screen in front of them–which, in any case, was the natural object to look at as it contained split times, stroke rates, and time information–to facilitate high-quality eye-tracking data. Once the rowers were ready, they completed each of the six conditions with the following order of proceedings:

1. Baseline recording with eye trackers, approximately 15 s.
2. Specific task instructions: Information about the upcoming condition, and a reminder to look at the screen in front of them throughout the trial
3. Rowing trial
4. Two-minute break
5. Open-ended question (which will not be the focus of this article)
6. Questionnaires: NASA-TLX as well as a self-made questionnaire addressing the rowers' thought process and focus while rowing (the latter scale will not be the focus of this article)

For the conditions with medium ML (race plan conditions), the rowers were also asked to state their race plan verbally after receiving the specific task instructions.

Before leaving, participants were debriefed and told to keep information about the mental manipulations in the study to themselves, so that the next participants would be naive and not given the chance to (mentally) prepare for the different ML conditions.

Data pre-processing and analyses

The code used for processing the eye-tracking and rowing ergometer data can be openly found at https://github.com/henrher/Rowing_EyeTracking.

Eye-tracking data extraction and filtering. First, output from the Pupil Labs recordings was down-sampled so that we had binocular data measured at 60Hz for both eye trackers.

For baseline trials, we extracted data from a period of 10 s where the participants sat still on the ergometer while looking at the screen in front of them. For rowing trials, we removed the first and last 10 s, so that we kept data from the 160 s in the middle of the trial to reduce noise (e.g., from the explosive movements that typically initiate a rowing trial).

To get a measure of blink rate, we counted the samples where pupil diameter equaled zero within a trial. However, if these zero values were less than 400ms apart in time, they were counted as part of the same blink (e.g., see Tanaka & Yamaoka, 1993).

Next, valid pupil size data were extracted. We used a stepwise filtering approach inspired by Bishop et al. (2021). The data were cleaned in order to keep valid pupil size data with realistic values captured during non-blink periods (Mathôt et al., 2018). Specifically, the following samples were filtered out, based on a script in R:

1. Samples where pupil diameter equaled zero

2. Remaining samples where the rate of change in pupil diameter, from one sample to the next, was greater than 2 standard deviations of the mean change rate.
3. Remaining samples where pupil diameter was either
    a. 3 standard deviations above the mean value
    b. 2 standard deviations below the mean value

Finally, a Savitzky-Golay filter (order = 3, window = 15) was added to smooth the data. The mean pupil size measured during the baseline was subtracted from the mean pupil size captured during rowing, to get a measure of pupil size change in each rowing condition. This difference value served as the dependent variable for pupil analyses.

To include data from a participant's eye, we set a cutoff of 50% valid pupil data during the rowing trials. For several rowers, data from only one eye met this criterion for all trials. Hence, we took the common approach of using data from one eye per participant, by keeping data from the "best-tracked eye". In most cases, both eyes could satisfy the criterion of 50% valid pupil data, and hence we used the participant's eye that had the greatest proportion of valid pupil data per rowing trial, on average. Two rowers (one elite and one non-elite) had less than 50% valid samples during certain rowing trials, and their eye-tracking data were therefore excluded from the analyses. In the remaining participants, the two eye-tracker systems had similar amounts of valid pupil data after filtering. Specifically, the SMI glasses' rowing trial recordings produced a mean of 80.37% (*SD* = 11.81%) valid pupil data, while the Pupil Labs glasses' rowing trial recordings resulted in a mean of 78.74% (*SD* = 15.69%) valid pupil data[4].

Rowing ergometer data. Split time for each stroke, providing a measure of rowing speed, was the key output for this current study. As with the eye-tracking data, the first and last 10 s of data were excluded. Next, we created a variable by subtracting participants' actual splits by their pre-planned split for each stroke in the given condition. This served as a manipulation check, to see how much faster or slower they rowed compared to what was intended. Given the fact that participants were instructed to maintain a constant, pre-planned split, we then converted the subtracted values into absolute numbers (i.e., by removing negative signs ahead of numbers) and calculated two performance variables:

1) mean stroke-by-stroke split deviation, namely how much their split deviated from the pre-planned split per stroke, on average

2) stroke-by-stroke split variability, namely the standard deviation of the discrepancy between their actual and pre-planned split

Data analyses. We conducted a series of mixed ANOVA analyses to explore the effect of ML and PL on the various measures of performance and mental effort. Skill level (elite and non-elite) was used as a between-subject fixed factor. A p-value of .05 was used as a significance cutoff for statistical comparisons. Greenhouse-Geisser corrected values were used in cases of sphericity violations. Partial eta squared (*ηp2*) was used as the main indicator of effect sizes, with .01, .06, and .14 indicating small, medium, and large effects respectively. Pairwise comparisons with Bonferroni adjustment were used to further investigate significant effects. JASP (https://jasp-stats.org/) was the software used for statistical analyses.

Results

Group characteristics comparisons

To ensure that our elite and non-elite group were comparable with regards to age and baseline eye-tracking measures, yet different with regards to rowing skill, we conducted a series of independent samples t-tests on these participant characteristics. Table 2 provides an overview of these characteristics across groups. Elites and non-elites were comparable with regards to age.

As expected, elites had significantly more rowing experience, faster expected 2K race times, and faster chosen pace (split) levels for low and high PL respectively, as compared to non-elites. With regards to eye tracking, the groups displayed similar numbers of blinks during the baseline. Mean pupil sizes during baseline, however, were larger in elites than non-elites. The latter difference was likely due to the mechanics and algorithms in the SMI glasses (used by all elites except one) as compared to the Pupil Labs glasses (see Appendix A for elaborations and further comparisons between the eye trackers).

Table 2

Descriptive Statistics and Comparisons of Participant Characteristics Across Groups

	Elite	**Non-elite**	
	Mean (SD)	*Mean (SD)*	*t*
Age (years)	29.67 (6.06)	29.89 (11.70)	-.051
Rowing experience (years)	14.89 (5.84)	4.33 (4.95)	4.136*
Expected 2K time (s)	358.33 (10.90)	409.22 (14.37)	-8.466*
Chosen split, low PL (s)	112.44 (4.00)	127.56 (4.59)	-7.447*
Chosen split, high PL (s)	103.44 (3.40)	117.44 (4.07)	-7.929*
Blinks during baseline	1.29 (1.99)	1.15 (1.74)	.383
Mean pupil size during baseline (mm)	4.76 (.80)	3.49 (.86)	7.504*

*$^*p < .001$

Performance

Rowing. The average split (time spent per 500 meters given the current pace) during low PL was 112.10 (*SD* = 3.77) s for elites and 126.86 (*SD* = 3.71) s for non-elites. During high PL, elites rowed with an average split of 103.11 (*SD* = 3.63) s while non-elites rowed with an average split of 117.42 (*SD* = 3.45) s. Table 3 shows descriptive statistics for the discrepancy between these actual split times, as measured by the RP3 ergometer, and pre-planned split times, as agreed upon before the start of trials, for each condition. Note that the information in Table 3 is based on true, uncorrected values (hence the negative mean values indicating that participants rowed with a lower split (i.e., faster) than the intended target on average). As a manipulation check of physical effort, we conclude that participants showed satisfactory adherence to their respective intensity levels.

Table 3

Deviations Between Observed Split Times and Target Split Times Per Condition, in Seconds Unit

		Low ML	**Med. ML**	**High ML**
Low PL	Mean	-.26	-.40	-.91
	Range	-2.33 - 1.04	-3.80 - 1.10	-4.72 - .77
High PL	Mean	-.01	-.30	-.23
	Min.	-2.52 - 1.29	-3.82 - 1.06	-2.27 - 1.11

A three-way mixed ANOVA was carried out with participants' mean absolute deviation scores as the dependent variable, ML (low, medium, and high) and PL (low and high) as the within-participants factors, and skill level (elite vs. non-elite) as the fixed between-participants factor. The interaction effects of ML and PL($F(1.503) = 1.104$, $p = .331$, $\eta p^2 = .065$), ML and skill ($F(2) = .363$, $p = .698$, $\eta p^2 = .022$), PL and skill ($F(1) = 2.757$, $p = .116$, $\eta p^2 = .147$), as well as all three factors ($F(1.503) = .956$, $p = .375$, $\eta p^2 = .056$), were nonsignificant. Results showed significant effects of ML ($F(2) = 8.963$, $p < .001$, $\eta p^2 = .359$) and PL ($F(1) = 4.865$, $p = .042$, $\eta p^2 = .233$). Pairwise comparisons with Bonferroni correction revealed that split deviations were significantly larger with high ML ($M = 1.58$ s, $SD = .87$ s) than low ML ($M = 1.07$ s, $SD = .52$ s) ($t = 4.232$, $p < .001$, $d = .998$), while medium ML ($M = 1.31$ s, $SD = .82$ s) did not differ from low ML ($t = 2.014$, $p = .157$, $d = .475$) and high ML ($t = 2.218$, $p = .101$, $d = .523$). As for the physical manipulations, on the other hand, split deviations were greater in low PL ($M = 1.45$ s, $SD = .86$ s) than high PL ($M = 1.18$ s, $SD = .66$ s). As expected, there was also a main effect of skill level ($F(1) = 6.733$, $p =.020$, $\eta p^2 = .296$), suggesting that elites ($M = 1.02$ s, $SD = .59$ s) deviated less from the pre-planned split target than non-elites ($M = 1.62$ s, $SD = .82$ s).

A similar three-way mixed ANOVA was conducted for the second performance measure, split variability (i.e., standard deviation of the absolute deviation values). Again, there were no significant interactions between ML and PL ($F(1.462) = .116$, $p = .828$, $\eta p^2 = .007$), ML and skill ($F(2) = .107$, $p = .899$, $\eta p^2 = .007$), PL and skill ($F(1) = 1.743$, $p = .205$, $\eta p^2 = .098$), or the three variables together ($F(1.462) = .697$, $p = .465$, $\eta p^2 = .042$). There were significant main effects of ML ($F(2) = 8.841$, $p < .001$, $\eta p^2 = .356$), PL ($F(1) = 12.412$, $p = .003$, $\eta p^2 = .437$), and skill level ($F(1) = 21.819$, $p < .001$, $\eta p^2 = .577$). Pairwise comparison with Bonferroni correction suggested significantly larger split variability in high ML ($M = 1.13$ s, $SD = .47$ s) as compared to low ML ($M = .78$ s, $SD = .31$ s) ($t = 4.190$, $p < .001$, $d = .988$), whereas medium ML ($M = .98$ s, $SD = .48$ s) did not significantly differ from low ML ($t = 2.403$, $p = .067$, $d = .566$) and high ML ($t = 1.787$, $p = .250$, $d = .421$). Participants had significantly higher split variability during low PL (M

= 1.09 s, *SD* = .51 s) than high PL (*M* = .84 s, *SD* = .34 s). The significant effect of skill level was about medium sized (ηp^2 = .577)–a notably larger effect than what was observed with split deviation as the dependent variable (ηp^2 = .296). Elites (*M* = .78 s, *SD* = .38 s) rowed with less split variability than non-elites (*M* = 1.14 s, *SD* = .44 s). Figure 1 displays split variability in elites and non-elites across ML.

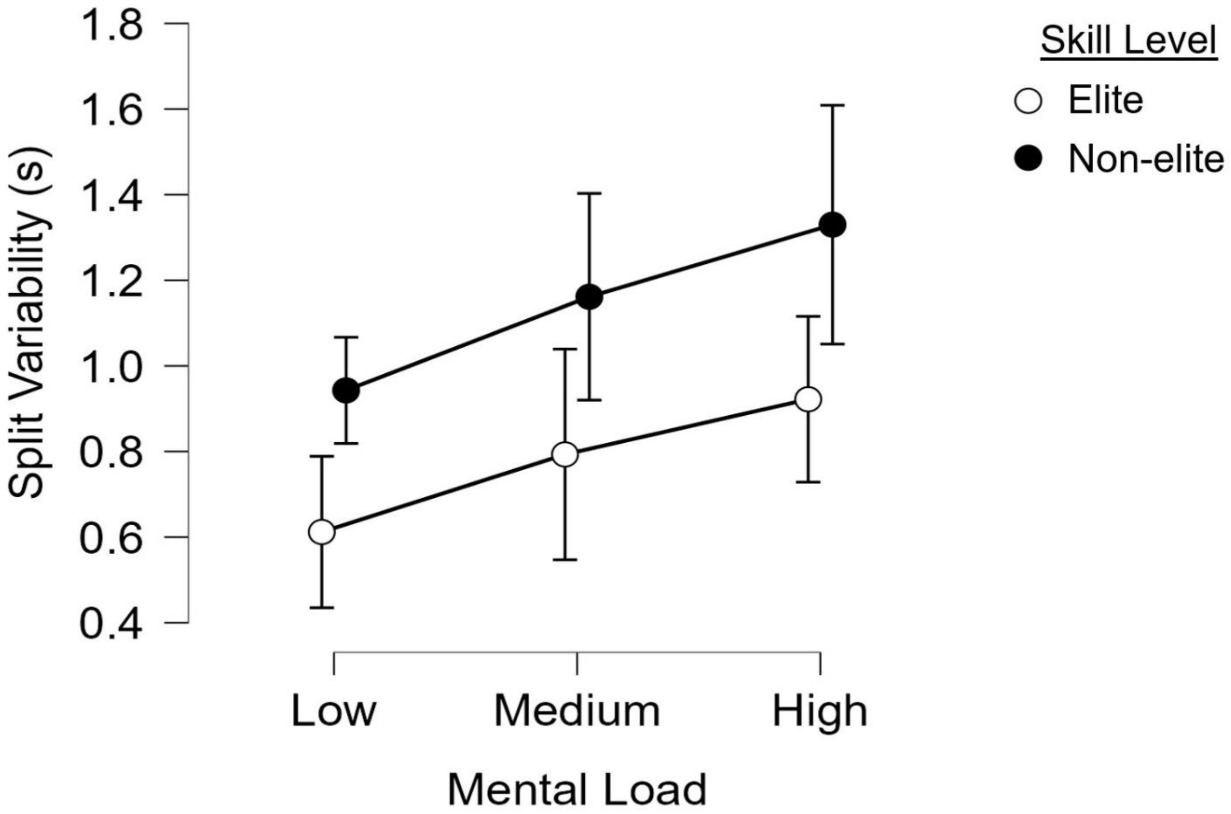

Figure 1 – Split variability across skill levels and mental load. Error bars indicate 95% confidence intervals.

Math. As a manipulation check, we report findings from participants' responses to math problems in the dual-task trials. Descriptive statistics suggested that all participants answered correctly to a minimum of five problems per condition, and the maximal score observed was 11 correct answers (out of 12 possible). A 2 x 2 mixed ANOVA suggested that more math problems were solved correctly during low PL (*M* = 7.89, *SD* = 1.64) than high PL (*M* = 7.06, *SD* = 1.47), *F*(1) = 7.563, *p* = .014, ηp^2 = .321. There was no significant effect of skill level (*F*(1) = .006, *p* = .937, $\eta p^2 < .001$) and no significant interaction between PL and skill (*F*(1) = .034, *p* = .857, ηp^2 = .002). These findings suggest that the math problems were equally challenging and attended to across the two skill levels.

Eye tracking

Blink rates. Descriptive statistics suggested individual variations in blink rates. Figure 2 provides an indication of the number of blinks, spread across trials. While the highest number of blinks in a single trial was 110, three (non-elite) participants finished rowing trials with zero blinks.

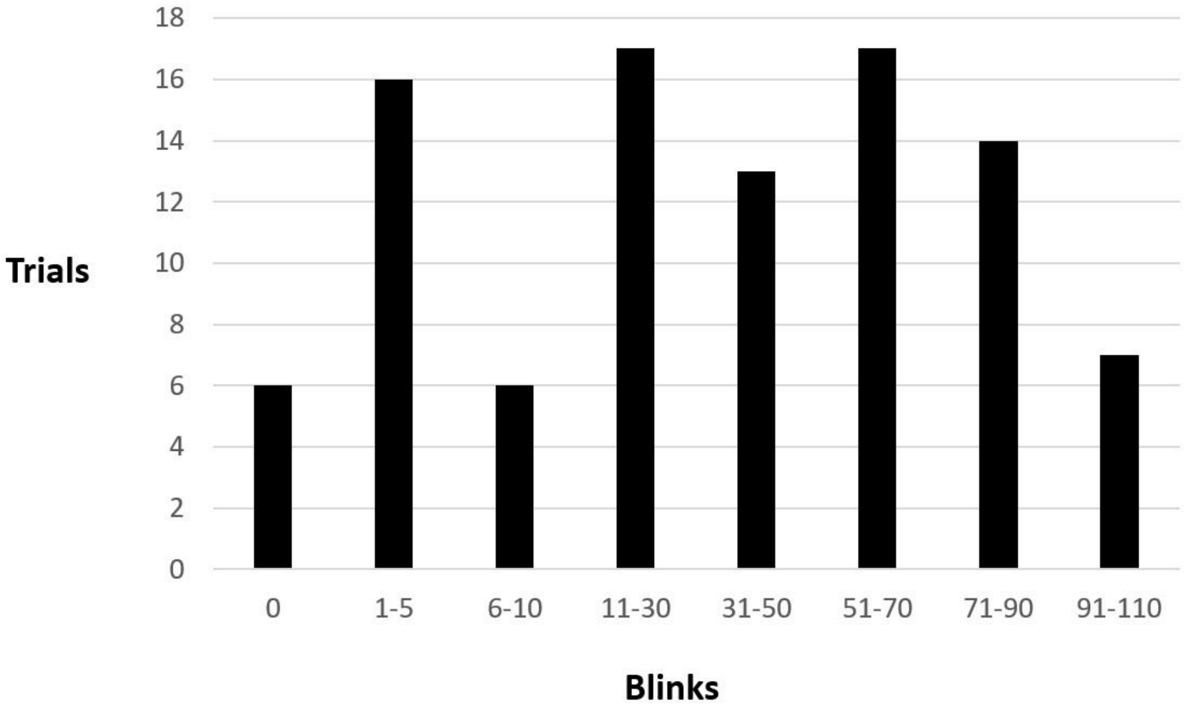

Figure 2 – The number of trials with different numbers of blinks.

A three-way mixed ANOVA revealed a significant interaction between ML and PL (*F*(2) = 3.638, *p* = .039, *ηp2* = .206). Yet, the pairwise comparisons revealed the same pattern of results for both physical intensities, namely that high ML involved significantly higher blink rates than the other two levels of ML (which were not different from each other), both during low and high PL. Descriptively, for low and medium ML, blinks were more frequent during high PL. For high ML (dual task), on the other hand, blinks were more frequent during low PL. Figure 3 displays blinks across ML and PL. The remaining interactions between ML and skill (*F*(1.104) = .221, *p* = .669, *ηp2* = .016), PL and skill (*F*(1) = .368, *p* = .554, *ηp2* = .026), and the three-way interaction (*F*(2) = 2.998, *p* = .066, *ηp2* = .176) were nonsignificant.

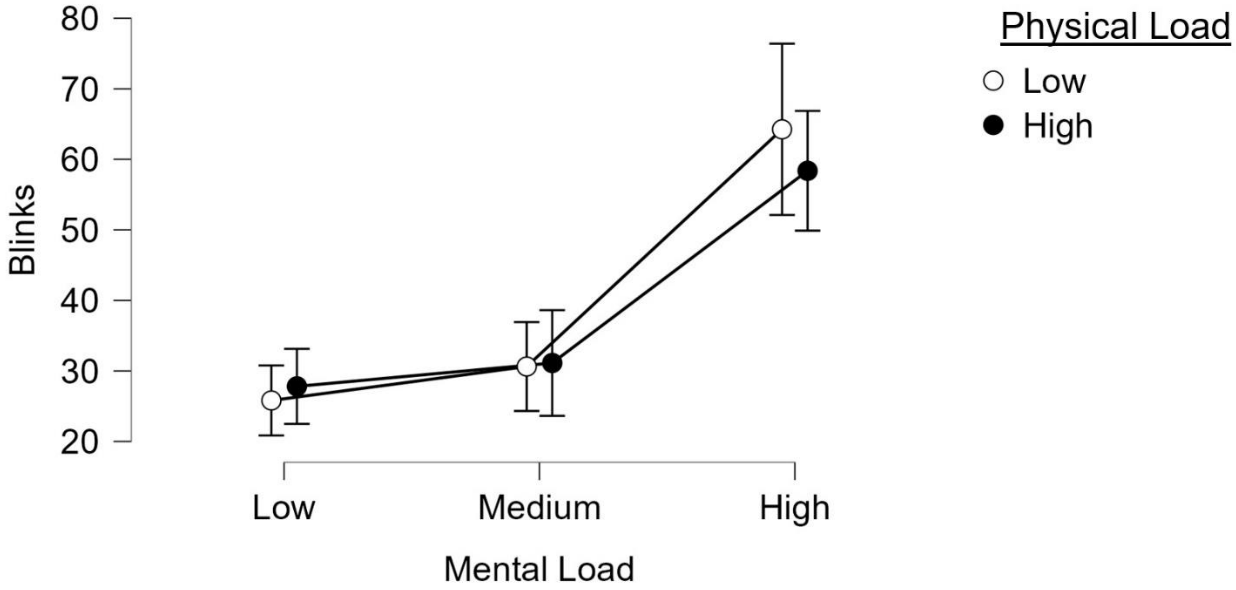

Figure 3 – Blink rates across load in all participants. Error bars indicate 95% confidence intervals.

The results further showed that participants' blink rates were significantly affected by ML, *F*(1.104) = 23.912, *p* < .001, *ηp2* = .631. Pairwise comparisons indicated that significantly more blinks occurred in the high ML, (*M* = 61.31, *SD* = 29.94) as compared to both the medium ML (*M* = 30.88, *SD* = 31.04) (*t* = 5.582, *p* < .001, *d* = 1.395) and low ML (*M* = 26.81, *SD* = 26.82) (*t* = 6.327, *p* < .001, *d* = 1.582), whereas the medium and low ML conditions were not significantly different (*t* = .745, *p* = 1.000, *d* = .186). No main effect of PL (*F*(1) = .319, *p* = .581, *ηp2* = .022) or skill level (*F*(1) = 1.815, *p* = .199, *ηp2* = .115) was found in relation to blink rates.

Pupil size change. Table 4 provides descriptive statistics for pupil dilations across conditions for participants included in eye-tracking analyses (*n* = 16).

Table 4

Pupil Dilation (mm) Across Conditions

	Low ML		**Medium ML**		**High ML**	
	Mean (SD)	*Range*	*Mean (SD)*	*Range*	*Mean (SD)*	*Range*
Low PL	.47 (.26)	.06 - .89	.52 (.28)	.20 - 1.26	.74 (.44)	-.02 - 1.68
High PL	.66 (.38)	.17 - 1.46	.57 (.36)	.13 - 1.51	.75 (.46)	.12 - 1.69

A three-way mixed ANOVA revealed a significant interaction between ML and PL (*F*(2) = 4.075, p = .028, ηp^2 = .225). Pairwise comparisons suggested that for the low ML conditions, pupil size changes were significantly greater with high PL (.66 mm, *SD* = .38 mm) than low PL (.47 mm, *SD* = .26 mm), *t* = 3.701, *p* = .010. This was also suggested by a plot showing pupil dilations across ML, with low and high physical intensities respectively–see Figure 4. Furthermore, when investigating pairwise comparisons of mental manipulations during low PL, it was evident that the high ML condition involved significantly larger pupil dilations as compared to both the low ML (*t* = 4.272, *p* = .001) and medium ML (*t* = 3.395, *p* = .021) conditions, with no significant difference between medium and low ML (*t* = .877, *p* = 1.000). However, during high PL, no pairwise comparison between ML levels reached statistical significance. That is, when rowing at the highest physical intensity in this study, the high ML condition did not involve pupil dilations that were significantly different from low ML (t = 1.525, p = 1.000) and medium ML (t = 2.930, p = .079) conditions, and the difference between medium and low ML was also nonsignificant (t = -1.405, p = 1.000). The remaining interaction results showed no significant effects of ML and skill ($F(2) = 2.756$, p = .081, $\eta p^2 = .164$), PL and skill ($F(1) = 1.584$, p = .229, $\eta p^2 = .102$), or the three-way interaction term ($F(2) = 1.287$, p = .292, $\eta p^2 = .084$) on pupil dilation.

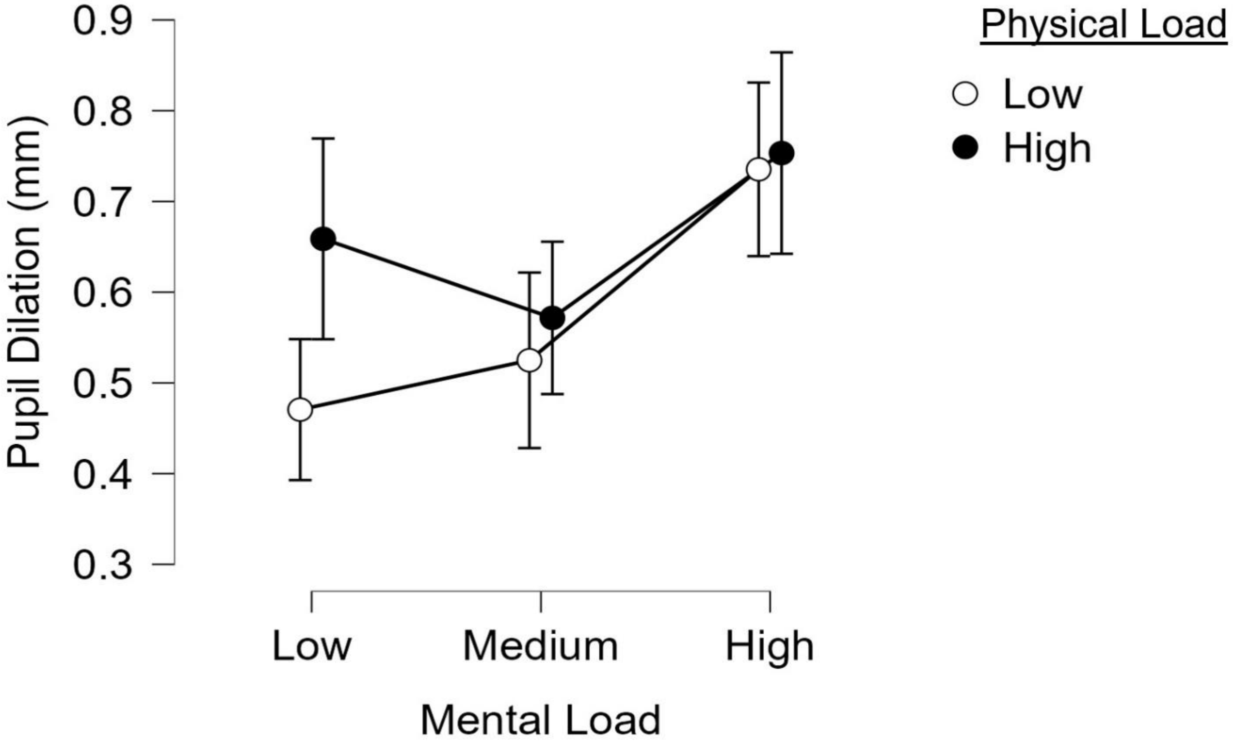

Figure 4 – Pupil dilations across load in all participants. Error bars indicate 95% confidence intervals.

The results further showed significant main effects of ML ($F(2) = 8.379$, p = .001, ηp^2 = .374) and PL ($F(1) = 5.792$, p = .030, $\eta p^2 = .293$). For ML, pairwise comparisons with

Bonferroni correction revealed that high ML led to significantly larger pupil dilations (*M* = .74 mm, *SD* = .44) than both medium ML (*M* = .55 mm, *SD* = .32 mm) (*t* = 3.689, *p* = .003, *d* = .922) and low ML (*M* = .56 mm, *SD* = .33 mm) (*t* = 3.381, *p* = .006, *d* = .845), while the latter two conditions were not significantly different (*t* = -.308, *p* = 1.000, *d* = .077). Further, the pupil dilated significantly more when PL was high (*M* = .66 mm, *SD* = .40 mm) than with low PL (*M* = .58 mm, *SD* = .35 mm).

A main effect of skill was also found, *F*(1) = 5.367, *p* = .036, ηp^2 = .277. Elites' pupils (*M* = .79 mm, *SD* = .41 mm) dilated substantially more than non-elites' pupils (*M* = .45 mm, *SD* = .25 mm). Figure 5 displays pupil dilations in elites and non-elites across ML.

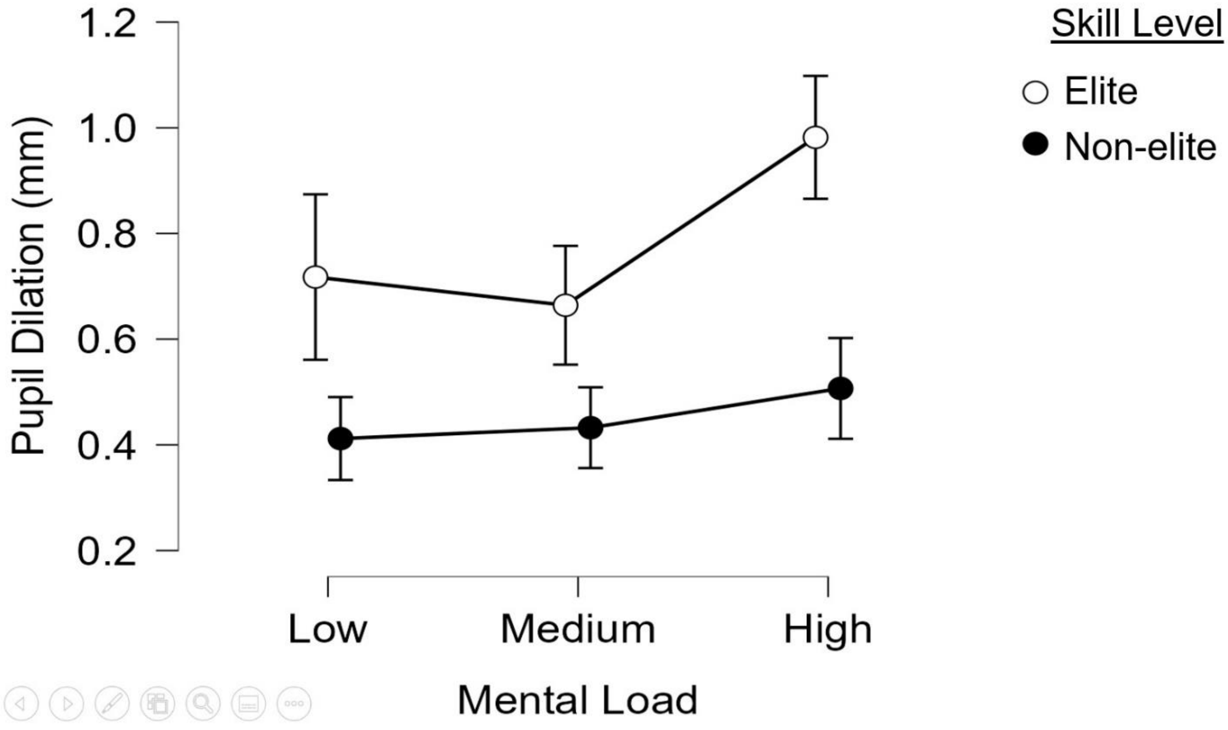

Figure 5 – Pupil dilations in elites and non-elites across mental load. Error bars indicate 95% confidence intervals.

Self-report data

Three-way mixed ANOVAs revealed no significant interaction–with all possible combinations of ML, PL, and skill level as the interaction terms–for any item of the NASA-TLX scale. Hence, the remainder of this section will elaborate upon the main effects of task load and skill level on questionnaire responses, item by item. Descriptive information is presented in the unit of percentages where 100% indicates the highest score possible, as rated on the visual analogue scale of the questionnaire.

A three-way mixed ANOVA revealed significant main effects of ML ($F(1.391) = 104.542$, $p < .001$, $\eta p^2 = .867$) and PL ($F(1) = 4.841$, $p = .043$, $\eta p^2 = .232$) on Mental demand ratings. Table 5 summarizes test statistics from pairwise comparisons following significant main effects of ML for all questionnaire items. Pairwise comparisons with Bonferroni adjustment revealed that Mental demand was rated increasingly higher from low (M = 29.94%, SD = 15.99%) to medium (M = 44.99%, SD = 17.96%) to high ML (M = 77.98%, SD = 14.19%), with significant differences ($p < .001$) between all levels. This partly served as a manipulation check, indicating that the ML had the intended effect. Participants also reported greater Mental demand in high PL (M = 53.38%, SD = 23.85%) than low PL (M = 48.56%, SD = 27.47%). Mental demand ratings were not significantly different across skill levels ($F(1) = 1.098$, $p = .310$, $\eta p^2 = .064$).

Table 5

Mental Load Comparisons with Bonferroni Adjustments for NASA-TLX Items

NASA-TLX item	Medium vs. Low ML		High vs. Low ML		High vs. Medium ML	
	t	*d*	*t*	*d*	*t*	*d*
Mental demand	4.427*	1.043	14.135*	3.332	9.708*	2.288
Temporal demand	.723	.171	6.483*	1.528	5.759*	1.357
Performance	.365	.086	-3.050*	.719	-3.415*	.805
Effort	2.188	.516	4.878*	1.150	2.690*	.634
Frustration	.026	.006	9.386*	2.212	9.360*	2.206

*$p < .05$

As another manipulation check, we noted that the next item, Physical demand, received ratings that were significantly affected by PL ($F(1) = 100.480$, $p < .001$, $\eta p^2 = .863$), with no effect of ML ($F(2) = .966$, $p = .391$, $\eta p^2 = .057$) or skill level ($F(1) = 1.584$, $p = .226$, $\eta p^2 = .090$). Physical demand received an average rating of 30.16% (*SD* = 15.91%) after rowing with low PL and 58.18% (*SD* = 11.55%) after high PL. This aligns well with the intention of letting participants row with low and moderate physical intensities, respectively.

Temporal demand ratings were significantly affected by ML ($F(1.143) = 25.238$, $p < .001$, $\eta p^2 = .612$) and PL ($F(1) = 6.270$, $p = .023$, $\eta p^2 = .282$), with no effect of skill ($F(1) = 3.266$, $p = .090$, $\eta p^2 = .170$). Pairwise comparisons showed that Temporal demand was rated significantly ($p < .001$) higher in the high ML (*M* = 57.80%, *SD* = 21.75%), as compared to both low ML (*M* = 26.51%, *SD* = 14.88%) and medium ML (*M* = 30.00%, *SD* = 16.19%), with no significant difference between the latter two ($p = 1.000$). Temporal demand was rated higher when PL was high (*M* = 41.27%, *SD* = 23.16%) as compared to low (*M* = 34.93%, *SD* = 21.77%).

Performance ratings were significantly affected by ML ($F(1.453) = 7.032$, $p = .008$, $\eta p^2 = .305$) and skill level ($F(1) = 11.590$, $p = .004$, $\eta p^2 = .420$), with no effect of PL ($F(1) = 1.388$, $p = .256$, $\eta p^2 = .080$). Pairwise comparisons with Bonferroni adjustments revealed that Performance was rated significantly lower in high ML (*M* = 45.19%, *SD* = 22.03%), as compared to both low ML (*M* = 58.56%, *SD* = 17.62%), $p = .014$, and medium ML (*M* = 60.17%, *SD* = 15.98%), $p = .005$, with no significant difference between the latter two loads, $p = 1.000$. Performance ratings were significantly higher in elites (*M* = 62.63%, *SD* = 18.74%) than non-elites (*M* = 46.65%, *SD* = 17.45%).

Effort was significantly affected by ML ($F(1.486) = 11.939$, $p < .001$, $\eta p^2 = .427$) and PL ($F(1) = 42.068$, $p < .001$, $\eta p^2 = .724$). Pairwise comparisons with Bonferroni adjustments revealed that Effort was greater in high ML (*M* = 67.86%, *SD* = 21.46%), as compared to both low ML (*M* = 47.39%, *SD* = 21.62%), $p < .001$, and medium ML (*M* = 56.57%, *SD* = 18.25%), $p = .034$, while low and medium ML were not significantly different, $p = .108$. Effort was also greater in high PL (*M* = 64.76%, *SD* = 15.70%) than low PL (*M* = 49.79%, *SD* = 24.79%). Skill level did not impact Effort ratings ($F(1) = .051$, $p = .823$, $\eta p^2 = .003$).

Lastly, Frustration was significantly affected by ML ($F(2) = 58.571$, $p < .001$, $\eta p^2 = .785$) and skill level ($F(1) = 8.590$, $p = .010$, $\eta p^2 = .349$), with no effect of PL ($F(1) = 2.237$, $p = .154$, $\eta p^2 = .123$). Pairwise comparisons suggested that reported Frustration was significantly higher ($p < .001$) in high ML conditions (*M* = 67.68%, *SD* = 19.79%), as compared to both low ML (*M* = 27.69%, *SD* = 22.65%) and medium ML (*M* = 27.80%, *SD* = 20.57%), with no significant difference between the latter two conditions ($p = 1.000$). Frustration was higher in non-elites (*M* = 49.21%, *SD* = 27.19%) than elites (*M* = 32.90%, *SD* = 26.91%).

Discussion

This study took a novel approach to investigating athletes' mental effort. Unlike most previous sport studies, an elite sample was recruited from the higher end of the performance spectrum. These experts were compared with athletes involved in the same sport, yet at a much lower level. Overall, the findings contradict the "zombie hypothesis", as elites displayed similar or even higher levels of mental effort, depending on the measure in question, as compared to non-elites. The multimethod approach was able to shed light on different nuances regarding mental effort. Findings further suggest that future physiological research investigating mental effort in sports may involve tasks of high physical demand and still have interpretable results, thanks to recent developments in non-invasive, mobile eye-tracking technologies.

Given the current experimental approach, the following sections will discuss findings in terms of the manipulations, namely how mental load (ML) and physical load (PL), and interactions between them, affected the results, before effects involving skill level are discussed. Finally, we discuss methodological issues, implications, and conclusions concerning this study.

The effects of load on mental effort and performance

Overall, rowers' mental effort was significantly affected by both ML and PL, and interactions between the two. Interestingly, an increase in PL–from low to moderate physical intensity–was associated with an increase in mental effort, measured both subjectively (i.e., via ratings of Mental demand in NASA-TLX) and objectively (i.e., via pupil size change). This indicates that there is a notable mental component to physical effort. Such a relationship is also suggested by research demonstrating that mental fatigue leads to suboptimal physical endurance (Van Cutsem et al., 2017; Zering et al., 2017). Increasing or maintaining physical effort relies on mental effort and–despite the prevalence of mind-body dualistic views in western societies (e.g., Gendle, 2016)–one cannot completely disentangle the two in sports contexts.

Unsurprisingly, ML had the strongest impact on mental effort, as measured by effect sizes from various measurements. Yet, as with previous studies (e.g., Harris et al., 2017), subjective and objective findings diverged to some extent. Specifically, subjective measures followed the hypothesized pattern, insofar as participants reported increasing amounts of Mental demand going from low to medium to high ML, respectively. Objective measures, on the other hand, suggested no difference in mental effort between the low and medium ML, while the high ML involved more effort, as measured by both blink rates and pupil dilation.

The diverging findings associated with mental effort during low and medium ML may be due to the nature of the conditions as well as subtle differences in the subjective versus objective measurements. Following a self-composed race plan (as done with medium ML) may have been subjectively interpreted as additional demands on rowing, as compared to only being asked to focus on speed (as done with low ML). This may have led participants to rate Mental demand higher in medium ML than in low ML conditions. However, while medium ML required some “extra steps” throughout the trial, these steps may not have been that difficult or effortful to implement. Unlike many past studies, we let participants choose their attentional cues and make their own performance plan. Thus, they were likely familiar with the task cues they chose as part of their plan. Mental representations of the technical elements they focused on, both during low and medium ML, were likely part of long-term memory formations. Using information that is readily available in long-term memory is less effortful than manipulating novel information in short-term memory (Kahneman, 1973). The latter type of processing was required in the high ML conditions when math problems were presented. This involved more time pressure and led to increased mental effort as predicted. In sum, we argue that the findings reflect real nuances, namely that using a race plan was indeed mentally demanding, per se, but it did not require an increase in mental effort due to the familiar nature of the specific demands.

To the best of our knowledge, this is the first study to combine different eye-tracking measures of mental effort in sports and to include blink rates in this venture. Current findings suggest that blinking can reflect mental effort, even while exerting substantial physical effort. Descriptively speaking, participants showed varying blink rates. Three of the rowers were measured for 160 seconds during rowing trials without a single blink. Interestingly, these participants were in the non-elite group. Additionally, numerous trials were finished with few (1-

5) recorded blinks. These results extend previous observations from non-sporting contexts showing that individuals can perform a task for more than a minute without blinking (Pei et al., 2021; Ponder & Kennedy, 1927), suggesting that infrequent blinking is not an exclusive hallmark of expertise in sports (see Pei et al., 2021; Red Bull, 2017). The fact that blink rates increased when solving math problems, as compared to single-task rowing trials, is in line with past research on blinking during increased (non-visual) cognitive demands (Magliacano et al., 2020; Marquart et al., 2015). Inhibited blinking may be part of a strategy to exploit visual information (Fogarty & Stern, 1989), while more frequent blinking can be a sign of “attentional disengagement from external stimuli” (Nakano, 2015, p. 54) to concentrate on internal cognitive work. In the current study, participants may have blinked less during single-task rowing trials in order to focus more intently on the visual information provided via the monitor (giving them relevant performance feedback). When receiving the added challenge of math problems in the dual-task conditions, less cognitive resources may have been devoted to rowing and visual feedback, and more to internal mental calculations, as evidenced by increased blink rates and rowing speed variability.

Pupil dilations suggested a specific interaction effect of mental and physical manipulations in our study. When ML was low, and the only instruction was to row with a certain speed, rowers’ pupils were significantly more dilated when this speed involved high PL than low PL. There are several possible explanations for this. One reason could be that rowers were the most motivated or excited in this particular condition, since the low ML gave them the most freedom to focus on whatever they wanted, and the high PL most closely mimicked the physical intensity they normally use in competition. Given this relative freedom and ecological validity, the condition may have motivated the rowers to invest mental energy at will. Such *autotelic* expenditure of effort would be in line with characteristics of the flow experience (Jackson & Csikszentmihalyi, 1999).

Further, the lack of significant difference in pupil dilations across ML when pupils were measured at high PL, could be due to a *ceiling effect*. That is, pupil dilation due to physical effort may have left less room for dilation due to mental effort (Hayashi et al., 2010). More pupillometry studies during physically strenuous tasks, combining different task demands, are needed to further enlighten the relationship between different influences on pupil size.

Finally, ergometer data suggested that rowers' speed control (i.e., performance) was significantly affected by both ML and PL, albeit in opposite directions. That is, increasing PL led to better performance while increasing ML led rowers to a decline, as measured by speed precision (i.e., the discrepancy between actual and pre-planned splits) and speed variability (i.e., standard deviations in actual splits). As for the mental manipulations, rowing speed was more consistent during low ML than high ML, as measured by both speed precision and speed variability. This makes sense, as speed control was the main element that participants were asked to focus on during low ML, while high ML involved focusing on math problems in addition to the current speed. The detrimental effect of high ML on speed control, as compared to low ML, was likely a distraction consequence of the dual-task environment (which will be discussed further below, in the context of skill level).

Mental effort in elites vs. non-elites

The most notable difference between groups came from pupil dilation findings. The direction of these findings was unexpected, given the fact that neural efficiency and hypofrontality have been associated with skilled performance in past psychophysiological studies (Filho et al., 2021). In the present study, pupil dilation, which is a global index of cortical activity (Just et al., 2003; Larsen & Waters, 2018), indicated significantly *more* mental effort in skilled performers. Specifically, elite rowers displayed pupil dilations that were larger than non-elites across all rowing trials. The easiest way to interpret this finding is to reject the zombie hypothesis. If experts perform more effortlessly, they should reveal smaller dilations than lower-level performers–the opposite of what we found.

The more uncertain interpretation aspect is to conclude what the greater pupil size changes mean more exactly, in terms of what caused elites to have larger pupil dilations. Pupil size measurement is, like most methods in psychology, not an *invariant* method (Richter & Slade, 2017). One finding, such as pupil dilation, can indicate several phenomena (such as thinking hard, taking drugs, or entering a dark room). One challenge, as noted by Kahneman (1973) and others, is to distinguish mental effort from other psychological phenomena involving *arousal*. However, it is worth noting that it may not be sensible, let alone possible, to disentangle the two, since mental effort will often be accompanied by arousal. If one is given a difficult math problem, for instance, one will not only notice that thought processes are occupied with the problem, but also that one's physiological responses and feelings indicate uncertainty or even fear (of potential failure). It could be that a brain network (involving the amygdala and other structures) detects uncertainty based on environmental cues, "signaling the need for the implementation of cognitive control" (Mushtaq et al., 2011, p.4) to deal with the problem. Kahneman's (1973) solution to this issue was not to disentangle mental effort from arousal, but rather to consider mental effort as a specific type of arousal, namely one that is related to intensive attention. In a controlled environment, it is possible to conclude that pupil dilations were likely caused by task manipulations and not merely, say, performance anxiety or other phenomena. The experimental setting and findings of the present study support the notion that elite athletes invested more mental effort, namely cognition-related arousal (Kahneman, 1973), directed at the task at hand. For example, the elites may have been more intensely focused on feedback, whether internal (from bodily cues) or external (from the ergometer monitor).

The idea that elites' performance depended on the intensive attention to rowing, is also supported by the fact that participants showed performance decrement in the dual-task conditions when their cognitive capacities were occupied with math problems while rowing. Elites and non-elites were similarly affected by mental load and the dual task specifically. This contradicts previous studies on higher-level athletes in dual-task environments (Beilock et al., 2002; Gray, 2004), and it could be because the secondary cognitive task in the present study was more challenging than some of those employed in past studies since we employed math problems that were of varying difficulty and based on past cognitive research. Despite the performance decrement across ML, elite athletes were able to perform better overall, according to both the ergometer data and subjective reports across conditions, as compared to non-elites. Overall, the main story thus seems to be that elite athletes invested relatively high levels of mental effort to maintain relatively high levels of performance. This resembles the Type 2 performance suggested by Bertollo et al. (2016), involving high effort and good performance, as opposed to the other type of good performance, Type 1, which is more relaxed and in line with the neural efficiency hypothesis (Cheng et al., 2017). Similarly, the results are also in line with *clutch* states (Swann et al., 2016, 2017), which may be prevalent when athletes face a challenge and invest effort to achieve a clear proximal goal. Given the fact that the current rowing trials were only three minutes long, elites may have been able to take a high-effort clutch approach to "make it happen" (Swann et al., 2016) throughout the trials.

Finally, subjective responses to various measures of NASA-TLX also contradicted the zombie hypothesis. Specifically, we found no effect of skill level on reported Mental demand, Physical demand, Temporal demand, or Effort. One may expect “zombie elites” to report low levels on all these measures. However, we did find that elites reported less Frustration than non-elites. Overall, these findings are interesting to consider in relation to the pupil dilation findings suggesting more intensive attention in elites. They may have invested more mental effort without experiencing it as more effortful, at least not in a negative sense. This, again, would suggest that effort can be experienced in different ways. However, NASA-TLX does not provide a full survey of emotional experiences and further studies may look more in-depth at the experience of mental effort in sports, for example by combining psychophysiological methods with qualitative interviews.

Methodological considerations

There are some methodological aspects that should be considered when interpreting the findings of this study. First, the sample size was limited and determined by the number of elite rowers from Team Norway’s Olympic training group. Second, due to the pupil measurements, we tested participants in a controlled laboratory environment. As compared to competitive rowing on open water, rowing ergometer performance involved lower (a) technical complexity (e.g., fewer degrees of freedom from a biomechanical perspective), (b) environmental complexity (e.g., not having to adapt to waves or competitors' tactical maneuvers), and (c) physical intensity (i.e., moderate as opposed to maximal intensity). On the one hand, the lower complexity of the ergometer performance would seem to invite more automatic performance. Indeed, it has been suggested that *closed* sports such as diving, where athletes are free from interference by competitors, should involve more automaticity than more *open* sports such as tennis, where athletes must constantly adapt to opponents' actions (Birch et al., 2019). In that sense, one could expect on-water rowing to require more mental effort, as it has more open-ended aspects than the ergometer rowing in the present study. On the other hand, the moderate physical intensities used in the present experiment may invite more conscious thinking and attentional capacity devoted to rowing and technical details, as compared to the maximal physical effort that is used during competition (Hutchinson & Tenenbaum, 2007). Given the fact that we were able to extract meaningful psychophysiological data from rowing at moderate physical intensities, future sports studies on mental effort may try to go even further up the intensity ladder and thus approach resemblance of real-world competitive scenarios. In any event, the current findings can be considered relevant to competitive performance from a technical perspective, since we used a dynamic rowing ergometer that mimics open-water rowing mechanics (Kleshnev, 2005).

With regards to pupil data, the fact that the two groups in our study were measured with mostly different eye-tracking systems, due to technical problems, was unfortunate and presented a methodological obstacle. Interpreting and comparing data from different equipment is a frequent, yet notable challenge in psychophysiological research (e.g., Stonnington et al., 2008). In the present context, the SMI glasses (used by all elites except one) could conceivably have had qualities that allowed for greater pupil dilation values, as compared to the Pupil Labs glasses (used by all non-elites). However, we find no indication of such a difference in our comparisons of the two systems (see Appendix A). There are past studies that have conducted more systematic comparisons of SMI and Pupil Labs glasses and found comparable characteristics on many measures, albeit some differences, but these have been focused on gaze data, not pupillometry (MacInnes et al., 2018; Niehorster et al., 2020). It is thus unclear how relevant these studies are in the present context. For example, Niehorster et al. (2020) found a substantial difference in data loss in SMI and Pupil Labs systems for gaze data, while we found similar amounts of data loss in SMI and Pupil Labs systems for pupil size data. The novelty of the current study, and the lack of relevant comparisons of eye-tracking systems, call for future eye-tracking studies in physically demanding sports to replicate or discredit our findings.

Finally, the manipulations of mental load are associated with both strengths and weaknesses. First, we decided to give participants the freedom to choose their foci and task cues, especially in the medium ML conditions where the rowers used a self-composed race plan. An alternative path would be to dictate what the athletes were supposed to focus on, which is typical in focus of attention research. However, asking athletes to focus on researcher-generated task cues can easily be criticized for low ecological validity and it may be particularly detrimental to performance at higher skills levels (Winkelman et al., 2017). Being familiar with one's task cues is a key to success (Maurer & Munzert, 2013), and to come up with functional task cues for all participants, from novices to Olympic medalists, was regarded as unrealistic.

The opposite applies to our high ML conditions, where more control was gained by asking participants to verbally respond to math problems while rowing, at the expense of ecological validity. This manipulation was included to make sure we had a level of ML that would likely present a cognitive challenge, and to make sure that participants actually partook in this challenge. On this note, it has been suggested that adding a verbal response to cognitive tasks can affect pupil sizes and blink rates, due to motor demands and not merely cognitive demands (Brych et al., 2021). However, verbalization has been found to have a rather small effect on pupil size, as compared to the effect of task difficulty (Kahneman et al., 1968), and not to affect blink rates (Brych et al., 2021). Additionally, the motor component of responding verbally to math problems was relatively small compared to the general physical demands in our study, in contrast to past studies on cognition outside of motor performance contexts. In sum, we interpret the indices of mental effort during high ML as evidence of, in fact, mental activity and not motor demands.

Implications and conclusions

For the last decades, mental effort has not had an unanimously favorable reputation, especially in elite sports. Performers have been frequently advised not to pay attention to task-relevant cues (e.g., Baurès et al., 2018) and even avoid using their prefrontal cortex (Beilock, 2011). The present study joins a recent line of research using eye tracking to demonstrate that sports tasks are, in fact, cognitively demanding (Campbell et al., 2019; Carnegie et al., 2020).

The current findings suggest that ergometer rowing–even when merely asked to follow a certain pace (low ML) or focus on a self-selected performance plan (medium ML), without any competitors or external interference–involves a measurable amount of mental effort, as indicated by the pupil sizes captured during rowing in comparison to pupil sizes captured during resting baseline periods. However, the mental effort captured while rowing in single-task conditions (low/medium ML) was small in magnitude, as compared to mental effort in dual-task conditions (high ML). This brings us closer to more specific answers regarding how effortful a task-related focus in sports can be, namely somewhere in between sitting still while looking at a screen (i.e., baseline) and performing while solving math problems. From an applied perspective, the advice of not spending mental effort at all seems unfounded. An alternative recommendation, directed at coaches and athletes themselves, would be to explore how mental effort can be ideally devoted to task cues, given each performer's unique abilities and context.

Eye tracking has become an increasingly popular tool to uncover selective aspects of attention. For example, a recent study became the first to use eye tracking in elite football to explore what players look at during actual match play (Aksum et al., 2020). Hence, this technology is likely to be employed in future sports studies and training contexts. We recommend that future scholars and practitioners keep using eye tracking to explore selective attention, while also adding an emphasis on intensive attention. For example, investigations of gaze strategies in sports could be supplemented by recordings of pupil sizes or blink rates, depending on contextual factors such as light conditions. It seems likely that such investigations can add value–for example by revealing how certain situations in a football game involve more intensive attention than others–beyond the mere exploration of what athletes look at. By combining selective *and* intensive aspects of attention, sport psychology could experience a “boost” similar to that achieved by emphasizing intensity in the investigation of physical performance (Schimpchen et al., 2016). For now, Kahneman’s (1973) original comments on how psychology traditionally has favored the study of selective attention, at the expense of intensive attention, may still ring true to some extent. Sports appears to be a well-suited arena for addressing this gap.

Endnotes

[1] Based on the present study, as well as personal communication with other eye-tracking researchers, we believe that sweat caused the technical problems.

[2] Before each medium ML condition, participants gave verbal descriptions of their self-composed race plans. After each rowing trial, the participants gave oral statements regarding what they were thinking of or attending to while rowing (see the Procedure part of the Method section). A full, systematic analysis of the verbal statements is beyond the scope of this article. However, as a manipulation check, it is worth noting some trends suggesting that the participants had different task-related foci in medium ML conditions (when asked to use a race plan while maintaining a constant split) as compared to low ML conditions (when only asked to maintain a constant split).

Specifically, after medium ML conditions, a majority of participants reported that they focused on certain cues at certain stages of the rowing trial. This was also suggested in the rowers' pre-trial race plan statements. Cues could, for example, be coupled with (a) general periods of the rowing trial, such as focusing on having a fast stroke rate in the beginning, before lowering the stroke rate; (b) specific time periods of the rowing trial, such as focusing on sequencing the stroke movement during the middle minute (i.e., between 60 s and 120 s of the rowing trial). All participants, except for one elite and one non-elite rower, gave race plan descriptions that involved such coupling between task cues and time periods in at least one of the medium ML conditions (i.e., with low and/or high physical load). In the oral statements following the low ML conditions, on the other hand, no participant reported such coupling. Another common feature in several race plans, for rowers of both skill levels, was the counting of strokes, typically by counting 10 strokes at a certain stage of the rowing trial, which is a common mental strategy used to deal with physical fatigue during rowing competitions.

Overall, it seemed that medium ML conditions involved attentional constraints, as compared to low ML conditions. That is, foci were typically linked to certain time periods or restricted to a certain number of strokes during medium ML, as part of their race plan, whereas the task cues focused on in the low ML conditions were reportedly not linked to specific parts of the trial. This seems like a partial, plausible explanation for why medium ML was subjectively experienced as more mentally demanding than low ML (see the Results section).

[3] Due to technical issues with the eye trackers, three participants (two elites, one non-elite) agreed to re-do one rowing trial each. In all three cases, this involved the low ML conditions (i.e., rowing at a constant pace). In two of the participants, the condition involved low PL, and they were able to complete this condition at the end of the trials. The third rower participated in the condition with low ML and high PL, before he had to come back on a later day to re-run all trials. Hence, fatigue should not be an issue in these trials. Further, since all three participants re-did the condition with low ML, which is the most common type of rowing on the ergometer, we do not suspect a training effect or benefit from having done this before. These last trials were therefore included in the analyses for these participants.

[4] The percentage of valid pupil data was calculated based on the relationship between valid pupil data after filtering and the expected number of samples in the raw data, given our sampling frequency (in line with Niehorster et al., 2020). Data loss was mostly due to the pupil data filtering, as described in our Methods section, in which we excluded samples with blinks and

unrealistic pupil values. However, certain SMI trials had fewer samples than expected in the raw data due to uneven sampling by the eye tracker, and this was also considered as data loss. We did not regard the latter data loss as problematic, for several reasons. First, the total amount of data loss (in the raw data and filtered data combined) never exceeded the exclusion criteria of 50%. Second, the maximum time interval between samples in the raw data was .067 seconds. Third, the raw data loss was evenly spread across the trial periods. In sum, we conclude that we had sufficient data from the SMI and Pupil Labs systems to conduct eye-tracking analyses on rowing trials.

Acknowledgements

[Omitted for blind review]

References

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with scholastic aptitude test scores. *Science*, *205*(4412), 1289–1292. https://doi.org/10.1126/science.472746

Aksum, K. M., Magnaguagno, L., Bjørndal, C. T., & Jordet, G. (2020). What do football players look at? An eye-tracking analysis of the visual fixations of players in 11 v 11 elite football match play. *Frontiers in Psychology*, *11*, 2624. https://doi.org/10.3389/fpsyg.2020.562995

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), 1-20. https://doi.org/10.1167/14.4.1

Barrett, B. T., Cruickshank, A. G., Flavell, J. C., Bennett, S. J., Buckley, J. G., Harris, J. M., & Scally, A. J. (2020). Faster visual reaction times in elite athletes are not linked to better gaze stability. *Scientific Reports*, *10*(1), 1-9. https://doi.org/10.1038/s41598-020-69975-z

Baurès, R., Balestra, M., Rosito, M., & VanRullen, R. (2018). The detrimental influence of attention on time-to-contact perception. *Attention, Perception, & Psychophysics*, *80*(6), 1591–1598. https://doi.org/10.3758/s13414-018-1523-x

Beilock, S. (2011). *Choke: What the secrets of the brain reveal about getting it right when you have to*. Free Press.

Beilock, S. L., Wierenga, S. A., & Carr, T. H. (2002). Expertise, attention, and memory in sensorimotor skill execution: Impact of novel task constraints on dual-task performance and episodic memory. *The Quarterly Journal of Experimental Psychology Section A*, *55*(4), 1211–1240. https://doi.org/10.1080/02724980244000170

Bertollo, M., Fronso, S. di, Filho, E., Conforto, S., Schmid, M., Bortoli, L., Comani, S., & Robazza, C. (2016). Proficient brain for optimal performance: The MAP model perspective. *PeerJ*, *4*, e2082. https://doi.org/10.7717/peerj.2082

Birch, J. E., Moe, V. F., & Breivik, G. (2019). Knowledge, consciousness, and sporting skills. In M. L. Cappuccio (Ed.), *Handbook of Embodied Cognition and Sport Psychology* (pp. 223-246). The MIT Press

Bishop, L., Jensenius, A. R., & Laeng, B. (2021). Musical and bodily predictors of mental effort in string quartet music: An ecological pupillometry study of performers and listeners. *Frontiers in Psychology*, *12*, 653021. https://doi.org/10.3389/fpsyg.2021.653021

Red Bull (2017, April 16). *Blink of an Eye: Aksel er verdens raskeste mann!*. https://www.redbull.com/no-no/blink-of-an-eye-aksel-er-verdens-raskeste-mann

Breivik, G. (2013). Zombie-like or superconscious? A phenomenological and conceptual analysis of consciousness in elite sport. *Journal of the Philosophy of Sport*, *40*(1), 85–106. https://doi.org/10.1080/00948705.2012.725890

Brych, M., Händel, B. F., Riechelmann, E., Pieczykolan, A., & Huestegge, L. (2021). Effects of vocal demands on pupil dilation. *Psychophysiology*, *58*(2), e13729. https://doi.org/10.1111/psyp.13729

Campbell, M. J., Moran, A. P., Bargary, N., Surmon, S., Bressan, L., & Kenny, I. C. (2019). Pupillometry during golf putting: A new window on the cognitive mechanisms underlying quiet eye. *Sport, Exercise, and Performance Psychology*, *8*(1), 53-62. https://doi.org/10.1037/spy0000148

Carnegie, E., Marchant, D., Towers, S., & Ellison, P. (2020). Beyond visual fixations and gaze behaviour. Using pupillometry to examine the mechanisms in the planning and motor performance of a golf putt. *Human Movement Science*, *71*, 102622. https://doi.org/10.1016/j.humov.2020.102622

Chavez, E. J. (2008). Flow in sport: A study of college athletes. *Imagination, Cognition and Personality*, *28*(1), 69–91. https://doi.org/10.2190/IC.28.1.f

Cheng, M.-Y., Wang, K.-P., Hung, C.-L., Tu, Y.-L., Huang, C.-J., Koester, D., Schack, T., & Hung, T.-M. (2017). Higher power of sensorimotor rhythm is associated with better performance in skilled air-pistol shooters. *Psychology of Sport and Exercise*, *32*, 47–53. https://doi.org/10.1016/j.psychsport.2017.05.007

Christensen, W., Sutton, J., & McIlwain, D. (2015). Putting pressure on theories of choking: Towards an expanded perspective on breakdown in skilled performance. *Phenomenology and the Cognitive Sciences*, *14*(2), 253–293. https://doi.org/10.1007/s11097-014-9395-6

Christensen, W., Sutton, J., & McIlwain, D. J. F. (2016). Cognition in skilled action: Meshed control and the varieties of skill experience. *Mind & Language*, *31*(1), 37–66. https://doi.org/10.1111/mila.12094

Csikszentmihalyi, M. (2000). *Beyond boredom and anxiety*. Jossey-Bass.

Del Percio, C., Rossini, P. M., Marzano, N., Iacoboni, M., Infarinato, F., Aschieri, P., Lino, A., Fiore, A., Toran, G., Babiloni, C., & Eusebi, F. (2008). Is there a "neural efficiency" in athletes? A high-resolution EEG study. *NeuroImage*, *42*(4), 1544–1553. https://doi.org/10.1016/j.neuroimage.2008.05.061

Dietrich, A. (2004). Neurocognitive mechanisms underlying the experience of flow. *Consciousness and Cognition*, *13*(4), 746-761. https://doi.org/10.1016/j.concog.2004.07.002

Filho, E., Dobersek, U., & Husselman, T. (2021). The role of neural efficiency, transient hypofrontality and neural proficiency in optimal performance in self-paced sports: A meta-analytic review. *Experimental Brain Research*, *239*, 1381-1393. https://doi.org/10.1007/s00221-021-06078-9

Fitts, P. M., & Posner, M. I. (1967). *Human Performance*. Brooks/Cole Publishing Company.

Fletcher, K., Neal, A., & Yeo, G. (2017). The effect of motor task precision on pupil diameter. *Applied Ergonomics*, *65*, 309–315. https://doi.org/10.1016/j.apergo.2017.07.010

Fogarty, C., & Stern, J. A. (1989). Eye movements and blinks: Their relationship to higher cognitive processes. *International Journal of Psychophysiology*, *8*(1), 35-42. https://doi.org/10.1016/0167-8760(89)90017-2

Gendle, M. H. (2016). The problem of dualism in modern western medicine. *Mens Sana Monographs*, *14*(1), 141–151. https://doi.org/10.4103/0973-1229.193074

Gray, R. (2004). Attending to the execution of a complex sensorimotor skill: Expertise differences, choking, and slumps. *Journal of Experimental Psychology: Applied*, *10*(1), 42-54. https://doi.org/10.1037/1076-898X.10.1.42

Harris, D. J., Vine, S. J., & Wilson, M. R. (2017). Is flow really effortless? The complex role of effortful attention. *Sport, Exercise, and Performance Psychology*, *6*(1), 103–114. https://doi.org/10.1037/spy0000083

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hayashi, N., Someya, N., & Fukuba, Y. (2010). Effect of intensity of dynamic exercise on pupil diameter in humans. *Journal of Physiological Anthropology*, *29*(3), 119–122. https://doi.org/10.2114/jpa2.29.119

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*(3611), 1190–1192. https://doi.org/10.1126/science.143.3611.1190

Hutchinson, J. C., & Tenenbaum, G. (2007). Attention focus during physical effort: The mediating role of task intensity. *Psychology of Sport and Exercise*, *8*(2), 233–245. https://doi.org/10.1016/j.psychsport.2006.03.006

Jackson, S. A. (1996). Toward a conceptual understanding of the flow experience in elite athletes. *Research Quarterly for Exercise and Sport*, *67*(1), 76–90. http://dx.doi.org/10.1080/02701367.1996.10607928

Jackson, S. A., & Csikszentmihalyi, M. (1999). *Flow in sports*. Human Kinetics.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *47*(2), 310–339. https://doi.org/10.1037/h0078820

Just, M. A., Carpenter, P. A., & Miyake, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science*, *4*(1–2), 56–88. https://doi.org/10.1080/14639220210159735

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Kahneman, D., Peavler, W. S., & Onuska, L. (1968). Effects of verbalization and incentive on the pupil response to mental activity. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *22*(3), 186–196. https://doi.org/10.1037/h0082759

Kleshnev, V. (2005). Comparison of on-water rowing with its simulation on Concept2 and Rowperfect machines. *International Symposium on Biomechanics in Sports, Conference Proceedings Archive*, *23*. https://ojs.ub.uni-konstanz.de/cpa/article/view/853

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27. https://doi.org/10.1177/1745691611427305

Larsen, R. S., & Waters, J. (2018). Neuromodulatory correlates of pupil dilation. *Frontiers in Neural Circuits*, *12*. https://doi.org/10.3389/fncir.2018.00021

Nakano, T. (2015). Blink-related dynamic switching between internal and external orienting networks while viewing videos. *Neuroscience Research*, *96,* 54-58. https://doi.org/10.1016/j.neures.2015.02.010

Magliacano, A., Fiorenza, S., Estraneo, A., & Trojano, L. (2020). Eye blink rate increases as a function of cognitive load during an auditory oddball paradigm. *Neuroscience Letters, 736,* 135293. https://doi.org/10.1016/j.neulet.2020.135293

Marquart, G., Cabrall, C., & de Winter, J. (2015). Review of eye-related measures of drivers' mental workload. *Procedia Manufacturing, 3,* 2854–2861. https://doi.org/10.1016/j.promfg.2015.07.783

Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods*, *50*(1), 94–106. https://doi.org/10.3758/s13428-017-1007-2

Maurer, H., & Munzert, J. (2013). Influence of attentional focus on skilled motor performance: Performance decrement under unfamiliar focus conditions. *Human Movement Science, 32*(4), 730–740. https://doi.org/10.1016/j.humov.2013.02.001

MacInnes, J. J., Iqbal, S., Pearson, J., & Johnson, E. N. (2018). Wearable eye-tracking for research: Automated dynamic gaze mapping and accuracy/precision comparisons across devices. *BioRxiv*, 299925. https://doi.org/10.1101/299925.

Mushtaq, F., Bland, A., & Schaefer, A. (2011). Uncertainty and cognitive control. *Frontiers in Psychology*, *2*. https://www.frontiersin.org/article/10.3389/fpsyg.2011.00249

Niehorster, D. C., Santini, T., Hessels, R. S., Hooge, I. T., Kasneci, E., & Nyström, M. (2020). The impact of slippage on the data quality of head-worn eye trackers. *Behavior Research Methods, 52*(3), 1140-1160. https://doi.org/10.3758/s13428-019-01307-0

Pei, X., Qi, X., Jiang, Y., Shen, X., Wang, A. L., Cao, Y., ... & Yu, Y. (2021). Sparsely wiring connectivity in the upper beta band characterizes the brains of top swimming athletes. *Frontiers in Psychology*, 2918. https://doi.org/10.3389/fpsyg.2021.661632

Perrey, S., & Besson, P. (2018). Studying brain activity in sports performance: Contributions and issues. In S. Marcora & M. Sarkar (Eds.), *Progress in Brain Research* (Vol. 240, pp. 247–267). Elsevier. https://doi.org/10.1016/bs.pbr.2018.07.004

Ponder, E., & Kennedy, W. P. (1927). On the act of blinking. *Quarterly Journal of Experimental Physiology*, *18*(2), 89-110. https://doi.org/10.1113/expphysiol.1927.sp000433

Richter, M., & Slade, K. (2017). Interpretation of physiological indicators of motivation: Caveats and recommendations. *International Journal of Psychophysiology, 119*, 4–10. https://doi.org/10.1016/j.ijpsycho.2017.04.007

Schimpchen, J., Skorski, S., Nopp, S., & Meyer, T. (2016). Are “classical” tests of repeated-sprint ability in football externally valid? A new approach to determine in-game sprinting behaviour in elite football players. *Journal of Sports Sciences*, *34*(6), 519–526. https://doi.org/10.1080/02640414.2015.1112023

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General*, *122*(2), 166–183. https://doi.org/10.1037/0096-3445.122.2.166

Simpson, T., Ellison, P., Marchant, D., & Carnegie, E. (2022). Effects of attentional strategies on novice dart throwing, Quiet Eye duration and pupillary responses. *Journal of Motor Behavior*, *54*(1), 1–13. https://doi.org/10.1080/00222895.2020.1869681

Stonnington, C. M., Tan, G., Klöppel, S., Chu, C., Draganski, B., Jack Jr, C. R., ... & Frackowiak, R. S. (2008). Interpreting scan data acquired from multiple scanners: a study with Alzheimer's disease. *Neuroimage, 39*(3), 1180-1185. https://doi.org/10.1016/j.neuroimage.2007.09.066

Swann, C., Crust, L., Jackman, P., Vella, S. A., Allen, M. S., & Keegan, R. (2017). Psychological states underlying excellent performance in sport: Toward an integrated model of flow and clutch states. *Journal of Applied Sport Psychology*, *29*(4), 375–401. https://doi.org/10.1080/10413200.2016.1272650

Swann, C., Keegan, R., Crust, L., & Piggott, D. (2016). Psychological states underlying excellent performance in professional golfers: “Letting it happen” vs. “making it happen.” *Psychology of Sport and Exercise*, *23*, 101–113. https://doi.org/10.1016/j.psychsport.2015.10.008

Swann, C., Moran, A., & Piggott, D. (2015). Defining elite athletes: Issues in the study of expert performance in sport psychology. *Psychology of Sport and Exercise*, *16*, 3–14. https://doi.org/10.1016/j.psychsport.2014.07.004

Tanaka, Y., & Yamaoka, K. (1993). Blink activity and task difficulty. *Perceptual and motor skills*, *77*(1), 55-66. https://doi.org/10.2466/pms.1993.77.1.55

Van Cutsem, J., Marcora, S., De Pauw, K., Bailey, S., Meeusen, R., & Roelands, B. (2017). The effects of mental fatigue on physical performance: A systematic review. *Sports Medicine*, *47*(8), 1569–1588. https://doi.org/10.1007/s40279-016-0672-0

Vickers, J. N. (2016). Origins and current issues in Quiet Eye research. *Current Issues in Sport Science, 1*(101), 1-11. https://doi.org/10.15203/CISS_2016.101

Whitehead, A. E., Jones, H. S., Williams, E. L., Dowling, C., Morley, D., Taylor, J. A., & Polman, R. C. (2019). Changes in cognition over a 16.1 km cycling time trial using Think Aloud protocol: Preliminary evidence. *International Journal of Sport and Exercise Psychology*, *17*(3), 266–274. https://doi.org/10.1080/1612197X.2017.1292302

Whitehead, A. E., Taylor, J. A., & Polman, R. C. J. (2015). Examination of the suitability of collecting in event cognitive processes using Think Aloud protocol in golf. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/article/10.3389/fpsyg.2015.01083

Winkelman, N., Clark, K., & Ryan, L. (2017). Experience level influences the effect of attentional focus on sprint performance. *Human Movement Science*, *52*, 84–95. https://doi.org/10.1016/j.humov.2017.01.012

Zarjam, P., Epps, J., & Lovell, N. H. (2012). Characterizing mental load in an arithmetic task using entropy-based features. In *11th International Conference on Information Science, Signal Processing and Their Applications (ISSPA)* (pp. 199–204). https://doi.org/10.1109/ISSPA.2012.6310545

Zénon, A., Sidibé, M., & Olivier, E. (2014). Pupil size variations correlate with physical effort perception. *Frontiers in Behavioral Neuroscience*, *8*, 286. https://doi.org/10.3389/fnbeh.2014.00286

Zering, J. C., Brown, D. M., Graham, J. D., & Bray, S. R. (2017). Cognitive control exertion leads to reductions in peak power output and as well as increased perceived exertion on a graded exercise test to exhaustion. *Journal of Sports Sciences*, *35*(18), 1799–1807. https://doi.org/10.1080/02640414.2016.1237777

Appendix A:

Pupil size data from SMI versus Pupil Labs

The elite and non-elite participants used mostly different eye trackers in the current study. Specifically, seven out of eight included elites used the SMI glasses, whereas all included non-elites used Pupil Labs glasses. Hence, any effect of expertise could conceivably be confounded with the eye-tracking systems. While previous research has compared eye trackers from SMI and Pupil Labs and found comparable attributes on several performance tests, these tests have not been focused on pupil size measurements (MacInnes et al., 2018; Niehorster et al., 2020). To explore the comparability between SMI and Pupil Labs data for the current purposes, we contacted the elite group, who had already participated using SMI glasses. Four of them volunteered to come in for a retest, approximately six months after their first participation. While being a limited comparison of the eye trackers' qualities, this test-retest was deemed useful to examine pupil size data across the two different systems.

In the retest, the four elites rowed with the same procedure as they had done on the previous test, in the same conditions, only with Pupil Labs glasses this time. Unfortunately, one rowing trial in the retest was not properly recorded due to technical difficulties. However, another elite participant had been able to finish one condition with SMI glasses, before technical issues arose and he had to come back to start the whole experiment anew with Pupil Labs glasses. He thus had finished one of the conditions with both eye-tracking systems, and this condition happened to be the same that was not recorded in the other retested rower (i.e., low mental load, high physical load; constant pace rowing at 85% of max), so these data were added to the current analysis.

In sum, we used four complete datasets, with trials from five elite rowers, to analyze baseline pupil measurements and pupil size change (average pupil size during rowing subtracted by average pupil size during baseline) data. Descriptive statistics, correlation (Pearson's *r*) and paired-samples t-tests were used to investigate pairs of data, with values from the same participant in the same condition, from two different eye trackers. Analyses were conducted in JASP, and $p < .05$ was considered statistically significant.

Results

Baseline measurements revealed significantly larger pupil sizes when participants were measured with SMI glasses (M = 4.65 mm, SD = .80 mm) than Pupil Labs glasses (M = 2.92 mm, SD = .61 mm), $t(23) = 12.031$, $p < .001$, $d = 2.456$. Thus, it appears that there is a difference in the eye trackers with regards to absolute pupil size measurements. For example, the SMI system may use cameras that are positioned closer to the participants' eyes, or it may use a different algorithm to compensate for distance between the cameras and the eyes, as compared to the Pupil Labs system.

The paired pupil size change values had a significant, positive correlation, $r = .694$, $p < .001$. The t-test revealed no difference between eye trackers, $t(23) = .375$, $p = .711$, $d = .077$. From a descriptive perspective, mean pupil dilations for the two eye trackers were highly similar: .81 (SD = .46) mm with SMI and .79 (SD = .36) mm with Pupil Labs. The maximum mean pupil dilation measured with SMI was 1.68 mm, similar to the maximum dilation measured with Pupil Labs, which was 1.59 mm, and these values were obtained from the same individual.

We also found that the SMI and Pupil Labs glasses showed similar amounts of data loss. Specifically, an independent samples t-test revealed no significant difference when comparing the percentage of valid pupil data from the two systems in the main rowing trials of our study (see our Method section for descriptive statistics), and a paired samples t-test suggested no significant difference in the amount of valid pupil data from the two systems in the test-retest rowing trials, $p > .05$.

Overall, the results suggest that the skill-related pupil dilation effect found in this study was, in fact, a skill level effect, and not an effect of two groups using different eye trackers.