Know your game, from in-real life experts to video game experts: discriminating in-real life experts from non-experts using blinks and EAR-derived features

Gianluca Guglielmo, Michal Klincewicz, Elisabeth Huis in ‘t Veld, and Pieter Spronck

Abstract— **Serious games are an effective method of reproducing aspects of the complex interplay between environments and stakeholders in business situations. In the game we describe here, The Sustainable Port, players experience what it is like to make decisions in such a complex environment. Their aim in the game is to grow the Port of Rotterdam while keeping economic growth in balance with sustainability goals. In this study, we assessed whether experienced Port of Rotterdam employees (PoR employees) show different psychophysiological patterns, and more specifically EAR-derived features, compared to students. We did this on the assumption that physiological patterns will tell us something about how people who are familiar with the environment of the Port of Rotterdam, more specifically Port of Rotterdam employees, make decisions compared to those lacking such familiarity.**

Our sample consisted of 28 PoR employees and 65 students, all of whom played The Sustainable Port game and had their faces recorded with a camera. The Eye Aspect Ratio was extracted from these recordings and then from those we extracted EAR-derived features. Our results show that Port of Rotterdam employees perform better than students and that the two groups are characterized by different physiological variations in their EAR-derived features. A logistic regression model used to identify PoR employees and students obtained an F1 score of 0.62, a PR AUC score of 0.64, and a ROC AUC score of 0.70. performing significantly above baseline suggesting the effectiveness of using EAR-derived features for this task. Our interpretation was further confirmed by a pseudo-R2 score used to evaluate the goodness of fit of a logistic regression model on the entire dataset. We found that Port of Rotterdam employees had a lower variation in blink rate per minute (blinks/m) and higher variation in the root mean square difference of successive difference in blinks (RMSSD), the consecutive difference between two continuous blinks. Moreover, this study shows that our methods were robust enough to negate the effects of confounders, such as biological sex and age that affect some other studies that analyze blinks.

Index Terms— **Blinks, Expertise, Machine Learning, Maritime Port, Serious Games, Sustainability**

I. INTRODUCTION

Video games are a widely spread medium of entertainment [1]. They often use mechanics that keep players engaged for hours, experimenting with different forms of agency, and familiarizing themselves with new environments [2]. This engagement can be harnessed for a variety of purposes. For example, L’Óreal and IBM used video games for recruitment purposes [3] while the Port of Rotterdam used games to make employees and stakeholders experience new business concepts [4]. These selected applications are just a few of many examples of the way in which games can, to a certain extent, imitate, engage, and inform players about matters important to business.

The Sustainable Port is a digitalized version of a board game with the same name developed by the Barn (https://thebarngames.com/?lang=en) [5] from Delft, Netherlands. This board game aims to simulate some of the dynamism of a busy maritime port and introduce complexities to that simulation based on the European Green Deal goals (https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en). Prior to this study, we found that port employees and students performed differently in this game and the patterns of their in-game decisions were directly impacted by their experience working for the Port of Rotterdam. From these results, we speculated that The Sustainable Port simulated an important aspect of the interplay between the environment and stakeholder interests that employees are familiar with. To further explore this possibility, it was hypothesized that employees may cognitively engage with the simulation differently compared to lay people and that this would be reflected in physiological variations connected to that engagement. Where previous studies have looked at differences between novices and experts in terms of gaze-fixation patterns [6] or heart-rate variability [7], we will want to focus on differences in Eye-Aspect Ratio-derived features (EAR-derived features; such as blinks/m or duration), a signal detected using a camera [8], tracking the eye lids movements. Previous studies already provide evidence that expert video game players experience low variation in blinks/m during a game session [8, 9]. Similar results were also found in real-life where expert surgeons show a higher blinks/m rate during the suturing phase of a microsurgery task suggesting that expertise may be connected to blinks at least in specific phases of a task [10]. However, up to date (July 2024) the previously mentioned is the only study involving expertise and blinks-related information. Furthermore, generally speaking, many studies focusing on blinks-related information did not control for the effect of confounders such as biological sex, age, and the differences of blinks at baseline which may affect the blinks and other blinks-related information collected during a task.

Expertise has been extensively studied in real life using sensors to extract heart rate variability information [11], brain activity information [12], or eye movements information [13]. However, no study has focused yet on exploring if expertise learned in real life also affects blinks/m and more generally blinks-related patterns in simulations such as serious games. Proving this point may be beneficial for the future use of serious games for education, training, and hiring new employees. For this reason, we extracted EAR-based features from the Eye Aspect Ratio. This has the benefit of being fully non-invasive and requiring only a webcam. The results obtained provide further evidence that EAR-derived features can be used to discriminate experts (in our case PoR employees) from non-experts (students) during gameplay, but also showcase a groundbreaking method and its potential use to evaluate other phenomena such as drowsiness [14], mental workload [15], or cognitive performance [16].

II. Related works

A. Blinks, Brain, and Expertise

Human beings make an average of 17 blinks per minute (blinks/m) with women typically blinking more than men [17, 18]. Previous studies show that blinks have a duration spanning between 50 and 500 ms [17, 18], and intervals between one blink and another are generally expected to be between 2 and 10 seconds [19] with a minimal interval as short as 100 ms [20]. Previous studies connected blinks/m with mental workload [13], performance [16], and fatigue [21]. In general, blinks/m are effective in tracking processes in the central nervous system, and dopaminergic activity in particular, but also processes that involve the pre-frontal cortex, the striatum, and the ventral tegmental area [22].

Blinks/m have also been connected to expertise in video games such as Tetris [8] and Hearthstone [9], as with expertise in real life [10]. A previous study showed that, for example, there is an interaction between expertise and suturing phases during a microsurgery task where experts seem to have higher blinks/m during some phases of the task [10]. Similar results were found during a Hearthstone tournament where expert Hearthstone players had higher blinks/m during the tournament independently from the opponent they were facing [9]. Importantly, these aforementioned studies have the shortcoming of not correcting for the baseline blinks/m that participants have at rest which may influence the results, and the effect found in these studies. Another study, this time correcting for the baseline at rest performing a subtraction between baseline at the blinks/m during the task [23, 24], shows that expert players of Tetris show a lower variation in blinks/m per minute both when looking at the first minute of gameplay but also when looking at the entire game session suggesting that experts players may have a higher blinks/m independently from the level they are playing [8].

However, not only blinks/m but other EAR-derived features may help distinguish skilled people from less skilled ones. Such features include, but are not limited to, the blinks' average duration, the average blink intervals, and the root mean square difference of successive differences in blinks (RMSSD). For example, changes in the blinks' average duration and the RMSSD were connected to physiological changes due to the task [25, 26]. More specifically, the RMSSD is supposed to increase, together with the average blinks' intervals, during visually demanding tasks [25] while the average blinks duration was reported to decrease [26]. Given this specification, we may assume these features may play a role as well in discriminating experts from non-experts assuming global differences, beyond blinks/m, in how they experience the task.

B. Previous Findings in The Sustainable Port

The Sustainable Port has already been used in a previous study [5]. This study found that both PoR employees and students who played this game agreed that the game can be used to inform non-experts about the complexity characterizing port environments when it comes to the decision-making processes connected to it. Furthermore, most PoR employees stated that they used the experience they obtained in real life at the Port of Rotterdam to play this game and that the game simulated the dynamics occurring in port environments (at least for what concerns the Port of Rotterdam). More interestingly, PoR employees scored higher than students in this game (after controlling for age, biological sex, video game habits, and board game habits) further suggesting a transfer of experience between what is learned in real-life environments and how people perform in a game. Therefore, given this evidence, differences between PoR employees and students may not only emerge on the level of the final score obtained but also when looking at their physiological variations, such as blinks and blinks-related information, and in-game decisions.

III. Methods

A. The Sustainable Port: How to play?

The Sustainable Port was originally a board game developed by The Barn, a game company based in Delft (Netherlands). This board version of the game was developed to start a discussion about green transition and to have multiple players involved and interacting. The version we used, developed by our research team at Tilburg University, was a single-player version instead that allows collecting data about the decisions made by the players through the game. Our digitalized game also allows for the extraction of physiological data collected through the computer webcams given that the player has his gaze and face oriented towards the screen. Furthermore, this game allows for the extraction of the decisions the player made throughout the rounds as an easily downloadable CSV file.

To play Sustainable Port, the players have to first read and learn the instructions of the game and then complete 10 rounds. The instructions of the game can be always accessed through the "instructions" button present in the game environment throughout the entire session. At the beginning of each round the player is informed about the objectives of the game, the predictions about the upcoming rounds of the game, and the new technology available (upgrades and facilities). Such objectives are always accessible throughout the entire game by pressing the button "objectives". Furthermore, during all the rounds of the game, the game has two indicators of performance the CO_2 emissions and the Added Value (revenues) given by the facilities and upgrades present in the Port. Such measures of performance, and the names assigned to them, were defined by the designers of the original board game. The final aim of the game is to first of all have a CO_2 <= 10 in order not to lose the game and second a score as high as possible given by the function Added value – CO_2 emissions.

As the players start to play the game, they will see three distinct areas (See Fig. 1). The first area represents the port where 12 spaces are located representing the facilities that the player decides to build in their port. At the beginning of the game (round 1) all players start with the same 11 facilities in their ports. Each facility contains information about the CO_2 emitted by the facility, its Added Value (revenues), how many rounds it takes to demolish it, its demolition cost, and the upgrades available. Upgrades can either reduce the CO_2 emissions of that specific facility or increase its added value. Every time a facility has its demolition (or construction) process started the game will continue to demolish (or build) such facility during the upcoming round unless the player stops it. The second area in the game is the information hub placed in the center of the Port. This area provides information about the current overall CO_2 emissions the added overall value of the port, and the money the players have available during the round they are playing (calculated using the following function: Added value/2+5; as defined in the original board game). Such money must be spent by the end of each round, nudging the players to carefully think about their decisions in order to optimize money spent and changes performed in their ports. The third area present in Sustainable Port is the "Market" where the new facilities that can be built are available.

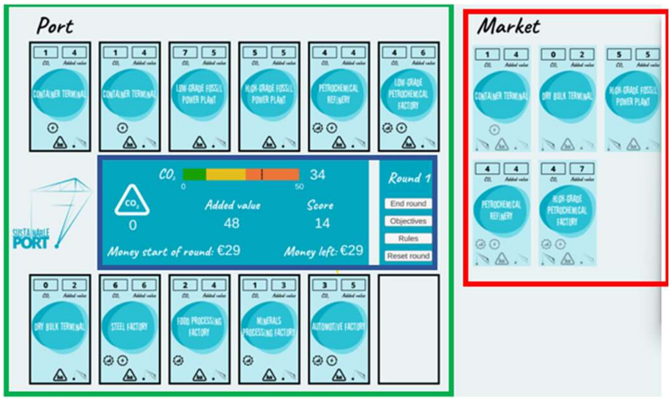


Fig. 1. An example of a round of The Sustainable Port. The Red rectangle contains the Market while the green one the port environment and the blue one the information about the performance of the player and general information about the round.

This game is conceived to have 3 phases, an introductory phase, a central phase, and a final phase. The introductory phase consists of rounds 1 and 2; during this phase, the player can familiarize with the game given that no new facility or upgrade is introduced. The second phase is composed of the rounds between round 3 and round 6 (included). This is the most crucial part of the game where the players can personalize their ports by building, demolishing, and upgrading. Finally, between round 7 and round 10, the final phase of the game occurs where the players can finalize their decisions and optimize their facilities through the use of the upgrades left.

B. Data Collection

B.1 Sample

In this experiment, a total of N = 109 people (Nfemales = 58, Nmales = 50, Nnd = 1; Mage = 27.27, SD = 11.20) were recruited to play The Sustainable Port (digital version). Seventy-five students were recruited through the recruitment system available at Tilburg University in exchange for formative credit. The student sample consisted of n = 27 males and n = 47 females (Nage = 21.20, SD = 3.4). One person did not declare their biological sex.

At the Port of Authority of the Port of Rotterdam, N = 34 participants were recruited from the HRM department, Strategy department, Finance department, Environmental management, Port Development department, Commercial Department & Policy Department of the Port Harbour Master. Such departments are relevant for our study since they are directly involved in the decisions affecting the future directions of the Port of Rotterdam. In this group, there were 9 Junior employees, 1 intern, and 24 senior employees. Overall, in this group, there were 23 males and 11 females with an average age of 40.67 (SD = 10.82). The participants working at the Port of Rotterdam were recruited using an anonymous Excel file shared through the Port of Rotterdam newsletter. The participants simply had to create an anonymous alphanumeric code with 10 digits without providing any information about their identity. The study was approved by the ethics committee of Tilburg University (REDC2021.35 + study number 3)

B.2 Procedure

At the beginning of the experiment, all the participants were asked to provide information about their biological sex, age, video game habits, and board game habits on a Likert scale ranging from 1 to 5 (Never, several times a year, several times a month, several times a week, every day) [5]. The participants working at the Port of Rotterdam were also asked about their level of seniority in the company and how many years they worked for the Port of Rotterdam. After gathering this information, participants were informed that video data collection commenced and were asked to watch a neutral video[1] from the OpenLAV library [27].

This video was repeated in a loop for 3 minutes and the recording was used to obtain the baselines for the EAR-derived features for each participant [8]. After watching the video, participants read the instructions of The Sustainable Port and played through the 10 rounds of the game. Generally, a

[1] The video used for this study is authored by Adrian Soare and it is available at the following link https://www.youtube.com/watch?v=dHG_eKFJHtM

complete experimental session lasts for approximately 60 minutes. At the end of the game, the participants could visualize if they reached the necessary CO_2 threshold not to lose the game (=< 10) and their final score. After having played the game, participants were asked when they developed confidence in the game mechanics (in terms of during which round) or if they did not develop confidence at all. The session ended after the participants answered some questions about their subjective experience with the game. Such results were reported in another study [5]. In our final sample, 13 participants, who did not develop confidence in the game mechanics were excluded. In this specific case, we asked the following question: "*During which round did you develop confidence in the game mechanics? (for example: which button or option is associated to specific actions)".* Therefore, these participants did not understand the mechanics intended here as the controls to play the game: where to click to perform the intended action. Consequently, it could not be assumed that these participants performed the actions they intended to do while playing the game. Out of these participants, 4 of them were PoR employees while 9 were students. From this study, we also excluded 2 PoR employee participants due to a corrupted video recording and 1 student who did not declare their biological sex, since our analyses required controlling for the biological sex of the participants. Eventually, the sample used in this study contained 93 participants.

The sample used for this experiment partly overlapped with the one reported in a previous study [5], but for this paper, we did not exclude participants who did not reach the threshold of CO_2 =< 10. This was done since we focused on discriminating Port employees from students and not on the score itself given the relatively small number of Port of Rotterdam employees taking part in the experiment. Overall, 6 students and 7 PoR employees did not reach the CO_2 threshold. However, we considered their game sessions serious attempts at playing the game since these players reported to have understood the game's mechanics. Nevertheless, before running an analysis focusing on blinks-based features, we used a multiple linear regression (controlling for age, biological sex, board game habits, and video game habits) to evaluate if the difference in score between the two groups can be found even when not considering the CO_2 threshold as in our current case. Table I and Table II show the information about the control variables, for the final sample (age, video game habits, board game habits, duration of the recording) we used in this study (for the analyses proposed in the next sections) evaluating respectively the differences between two groups (PoR employees and students) and the two biological sexes. This information provides preliminary insights about the nature of the sample collected. The p-values provided in Table I and Table II are based on Welch t-tests.

TABLE I
THE DIFFERENCES IN CONTROL VARIABLES FOR POR EMPLOYEES AND STUDENTS

	PoR employees (N = 28)	Students (N = 65)	*t*	*p*
Age	39.89 (SD = 10.83)	21.2 (SD = 3.46)	F(1,29.35) = 8.77	< .001***
Video game habits	1.86 (SD = 0.79)	2.85 (SD = 1.21)	F(1,75.48) = 4.62	< .001***
Board game habits	2.25 (SD = 0.57)	2.31 (SD = 0.58)	F(1,51.57) = 0.55	.58
Duration of the recording (min)	23.88 (SD = 6.47)	18.38 (SD = 5.83)	F(1,46.30) = 3.81	< .001***

*** Refers to p-values below .001.

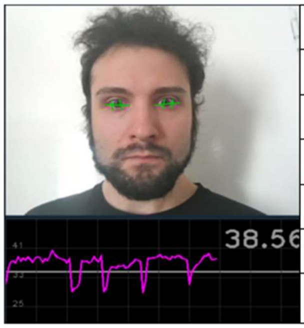
TABLE II
THE DIFFERENCES IN CONTROL VARIABLES FOR POR MALES AND FEMALES

	Males (N = 47)	Females (N = 46)	*t*	*p*
Age	30.06 (SD = 13.03)	23.52 (SD = 6.41)	*F*(1, 67.24) = 3.04	< .01**
Video games habits	2.74 (SD = 1.31)	2.35 (SD = 1.01)	*F*(1, 90.95) = 1.12	.27
Board games habits	2.23 (SD = 0.59)	2.37 (SD = 0.57)	*F*(1, 86.06) =1.62	.11
Duration of the recording (min)	20.35 (SD = 6.88)	19.72 (SD = 6.14)	*F*(1, 90.26) = 0.46	.65

** Refers to p-values below .01.

B.4 Eye Aspect Ratio Extraction

Similar to what was done in previous studies [8], the first step to detect blinks from video is to determine how to calculate the distance between the eyelids. Blinks occur when the eyelids close and therefore when the distance between each other is minimized. In order to keep track of the distance between the eyelids, landmarks P1-6 are projected onto the area between the eyelids, and the EAR formula is applied to the values of landmarks: (|P2-P6|+|P3-P5|)/(2|P1-P4|); the EAR signal is the average across the eyes. Landmarks are projected onto the eye using the *cvzone* library (Python), which uses the FaceMeshDetector detector function to process a video as in Fig 2.



Points	Left eye	Right eye
P1	243	385
P2	22	252
P3	24	254
P4	130	463
P5	160	387
P6	158	359

Fig. 2. The landmarks of the FaceMeshDetector and the corresponding EAR points for the left and right eyes. The peaks in the signal represent blinks in the EAR signal.

A value representing the average distance between eyelids typically ranges between 25 and 40 [8]. Every time a blink occurs, a peak is detected, as can be seen in the dips in the recording (Fig. 2). A stream of EAR values per frame of the entire recording can be saved as a .csv file.

B.4 EAR-derived features extraction

As shown in Fig. 2, we can easily see that peaks in the signal represent blinks. However, we still need to extract blinks, represented by sets of frames appearing as peaks on the EAR signal, and other EAR-derived features. In this study, we used the Isolation Forest with filtering proposed in a previous study [8]. Such a method detects blinks using an informed Isolation Forest and further functions to reduce the presence of noise. Furthermore, this method was tested on 30 fps recordings like ours. First of all, this Isolation Forest has a contamination parameter, defining the percentage of expected outliers, based on a rolling median of 100 frames with a median absolute deviation of 2.5 [28, 29] to determine the percentage of outliers and therefore the contamination parameter. Second, the Isolation Forest marks as outliers all of those sets of frames lasting between 2 frames and 15 frames which approximated the lower and upper bounds of blink duration (50 and 500 milliseconds) [19, 20]. During this step of the process, we obtained the candidate sets of outliers that are potential blinks. However, at this stage, a filtering function is implemented where sets of outliers with a distance lower than 4 frames (representing the minimal interval of at least 100 ms between one blink and another suggested by previous studies [20]) are detected. The frames missing between these two sets of outliers are then marked as outliers as well since they are considered outliers that the Isolation Forest failed to detect given the minimum blinks interval of 100 ms [20]. Finally, a last filtering function detects the final groups of outliers (blinks) after having controlled that their final length is between two frames and 15 frames. This process should exhaustively capture blinks that appear as peaks in the EAR signal.

The Isolation Forest with filtering in this previous study not only allows for the extraction of blinks and therefore blinks/m but also for other features such as the average distance between one blink and another (blinks intervals), the blinks' duration, and the RMSSD between one blink and another [8]. For this study, we focus on four main features: the blinks/m, the average blinks' duration, the average blinks' interval, and the RMSSD. We used these features since they may provide a more exhaustive overview of what occurs in blink patterns. For example, as mentioned in a previous section of this work, other features besides blinks/m, such as blink duration and RMSSD seem to be connected with phenomena such as visual workload and the allocation of visual resources [25, 26] that may be mediated by expertise.

B.5 Analyses pipeline and assumptions

Given the aim of this paper to both propose a methodology to evaluate EAR-derived features and to investigate if the extracted EAR-derived features can be used to discriminate PoR employees from students playing Sustainable Port this section provides an overview of the analyses run and their purpose. First of all, we evaluated if the patterns found in [5] concerning the differences in scores between PoR employees and students can be found in our sample as well (subsection: PoR employees and students score comparison). Second, we ran analyses aiming to evaluate the effect of confounders (such as age or the duration of the recording) on the EAR-derived features (subsection: methodological analyses). Such analysis was run since most of the studies focusing on blinks did not evaluate the confounders (age, and biological sex) that are present independently from the field of study [9, 10]. Third, we will evaluate if the baseline-corrected EAR-derived features (after having controlled for the effect of age, and biological sex on these features) can be used to discriminate PoR employees from students (subsection: PoR employees and students EAR-derived features' analyses).

B.5.1 PoR employees and students score comparison

This first analysis was run to evaluate if our study, independently from participants reaching the CO_2 threshold, presents the same patterns as [5] where PoR employees performed better on the game than students. For this reason, a multiple linear regression was run with the final score as the dependent variable and group (PoR employee, student), age, biological sex, board game habits, and video game habits as control variables. This approach was adopted to make sure to find the same patterns in our study (for what concerns PoR employees performing better than students) as it was found in a previous study using Sustainable Port [5].

B.5.2 Methodological Analysis

For the second set of analyses, we ran an ANOVA analysis comparing the two groups, using age as a covariate, to detect if there were baseline differences in baseline blinks/m, baseline average blinks duration, baseline average interval duration, and baseline RMSSD. Age was added to the analysis since the two samples have 2 different ages and previous studies suggest an effect of age on blink patterns where older women, for example, tend to blink more than younger ones [30]. This analysis was run to exclude baseline differences between the two groups. Then, for all four features separately, a mixed linear model was run to evaluate period (baseline vs task period) as within subject factor and, age, biological sex, and duration of the recording as between subject factor. In this case, we decided to use a linear mixed model given their robustness to non-normal residuals and heteroskedasticity that we found in the data [31]. To these variables, the interaction of age and biological sex was added as a term given previous study suggesting the effect of age [30, 41] and biological sex [8] on EAR-derived features.

After that, we ran an ANOVA on the baseline corrected measures obtained by subtracting the measures collected during the tasks from the baseline as similarly done in other studies [8, 20, 21] thus obtaining the baseline-corrected blinks/m, the baseline-corrected average blinks duration, the baseline-corrected average blinks interval, and the baseline-corrected average RMSSD (similar to was is generally done in heart rate variability studies [32]). In this specific analysis, we used the same predictors as in the linear mixed models (biological sex, age, duration of recording, and interaction between biological sex and age) to evaluate if the baseline corrected measures are still affected by these confounders. As in the case of baseline vs task measurements, we found often non-homogenous variance and non-normal residuals in our data both in the baseline measures and in the baseline-corrected measures. For this reason, we used the HC3 correction to obtain more reliable results as suggested in a previous study where HC3 was deemed to be robust both against non-normality of the residuals [33] but also performs well on non-homogenous variance-affected data in small imbalanced samples (< 250) like ours [34]. Furthermore, further evidence of the robustness of the HC3 correction is given by other studies suggesting that HC3 should be routinely used given also the low power that heteroskedasticity tests have on small samples [35]. Furthermore, this method should be also robust to the differences in age we found in our sample between males and females (see Table I). For this reason, in this study, we used type 2 ANOVAs implemented with an HC3 correction.

B.5.3 PoR employees and students EAR-derived features' analyses

These analyses aim to evaluate if baseline-corrected EAR-derived features can be used to discriminate PoR employees and students. We used logistic regression with stratified 5-fold cross-validation to evaluate if the four baseline-corrected features can be used to discriminate PoR employees from students after having balanced the sample weights. The metrics used for this purpose were the weighted F1 score, the area under the precision-recall curve (PR AUC), and the ROC AUC. Such metrics should provide reliable indicators with strongly imbalanced samples like ours (where PoR employees account only for 30% of the data) [36, 37, 38]. The baseline scores were defined using a dummy classifier implemented with the "stratified" strategy for the PR AUC and the "constant" strategy with the underrepresented class (PoR employees) for the F1 score as suggested in sklearn[2]. Given the imbalance in the data, for this analysis, we balanced the class weights in our logistic regression and evaluated the model on a weighted F1 score and a manually implemented PR AUC with balanced weights (given the lack of a function to calculate this score in sklearn). Finally, we also fitted the data using a logistic regression implemented with a pseudo-R2 (McFadden pseudo-R2) to evaluate which features are significant and to evaluate the coefficients and p-values of our features when fitting the data. Logistic regression was chosen for this analysis since our dataset, considering the four features selected, fits the suggested requirement of at least 5 events per variable in the underrepresented class (PoR participant N = 27) [34]. Furthermore, we evaluate the presence of multicollinearity by checking the variance inflation factor of our features where none of the features had a value higher than 10 suggested as the limit for multicollinearity in previous studies [39].

IV. Results

A. PoR employees and students' comparison

The results of this analysis, using multiple linear regression, suggest that similar patterns of significance to what found in [5] can be found when not considering the CO_2 <= 10 threshold needed not to lose the game ($F(5, 87) = 5.22, p < .001, R2 = .23$). More specifically, as showed in Table III, video games habits have a positive significant effect on the score while PoR employees (M = 35.11, SD = 13.28) scored significantly higher than students (M = 31.15, SD = 16.33). Furthermore, no difference was found between males (M = 36.47, SD = 14.25) and females (M = 27.85, SD = 15.68), and no effect of age was found. Table III shows the results of the multiple linear regression.

TABLE III
The Results of the multiple linear regression on the score

	B	**SD**	**U**	**L**	**t**	***p***
Biological Sex (Ref. female)	6.17	3.33	12.80	-0.45	1.85	.067
Age	-0.42	0.23	0.036	-0.89	-1.83	.07
Participants (Ref. students)	13.56	5.49	24.47	2.64	2.47	.016*
Digital game habits	4.12	1.44	6.99	1.25	2.85	.005**
Board game habits	2.99	2.62	8.19	-2.20	1.14	.257

* Refers to p-values < .05, while ** refers to p-values < .01.

[2] https://scikit-learn.org/stable/modules/model_evaluation.html

B. Method Selection

B.1 Baseline differences between groups

In this first step of our analyses, we evaluated the presence of baseline differences between PoR employees and students (controlling for the effect of age as a covariate) for the chosen features; in this case, we did not control for the duration of the recording since it was of 3 minutes for all the participants. This analysis was carried out to exclude the presence of significant findings in the next analyses due to baseline differences. The descriptive statistics of the features used can be found in Table IV.

TABLE IV
TABLE ILLUSTRATION THE FOUR EAR-DERIVED FEATURES FOR POR EMPLOYEES AND STUDENTS AT BASELINE

	PoR employees	Students
Blinks/m	M = 15.27 (SD = 9.34)	19.67 (SD = 8.95)
Average Blinks' duration (in milliseconds)	M = 171.60 (SD = 28.98)	M = 175.96 (SD = 27.11)
Average Blinks' Interval (in seconds)	M = 6.93 (SD = 8.74)	M = 3.79 (SD = 2.90)
Average RMSSD (in seconds)	M = 12.09 (SD = 17.57)	M = 6.41 (SD = 7.82)

Overall, in this analysis, we found no effects of the predictors. There was no effect on the blinks/m determined by group ($F(1,90) = 0.21$, $p = .65$) or the age of the participant ($F(1,90) = 1.31$, $p = .26$). Similar results were found in the other three chosen features. For the case of the RMSSD, there was no significant difference due to age ($F(1,90) = 1.68$, $p = .20$) and group ($F(1,90) = 0.75$, $p = .39$). Similarly, the average blinks' interval was not determined by age ($F(1,90) = 1.49$, $p = .23$) and group ($F(1,90) = 0.34$, $p = .56$). Finally, non-significant effects were found in the average blinks duration at baseline where neither age ($F(1,90) = 0.021$, $p = .89$) nor group ($F(1,90) = 0.25$, $p = .62$) are significant predictors.

B.2 Blink behavior during the baseline and task (Linear Mixed Models)

In this second step of our methodological pipeline, we wanted to evaluate the difference between baseline measures and measures collected during the task (for all the four features selected) controlling for variables such as age, biological sex, duration of the recording, and the interaction between biological sex and age. These analyses follow the example provided in a previous study [8] and want to appraise the effect of baseline measures on the measures obtained during the task independently from the main groups' analysis. Fig. 3 provides a visual representation of the data during the baseline and the task (The Sustainable Port game session) period for males and females similar to what was done in [8].

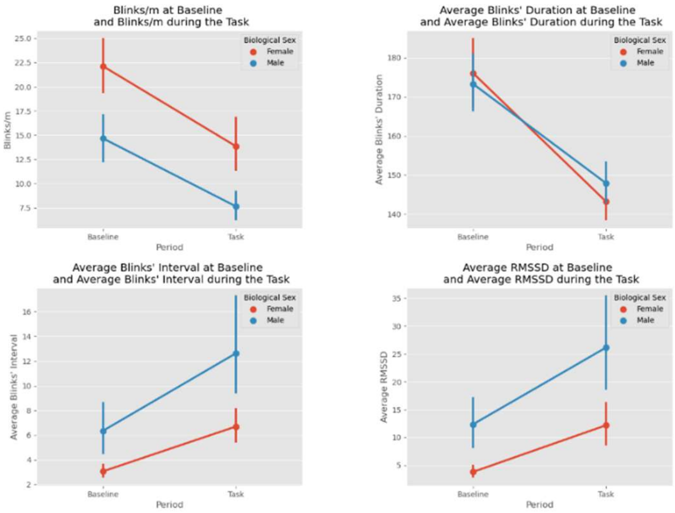


Fig. 3. The variation for the four EAR-derived features according to period and for the two biological sexes.

The results for the linear mixed models analysis can be found in Table V for the blinks/m, Table VI for the average blinks' duration, Table VII for the average blinks' interval, and Table VIII for the average RMSSD. For all the tables mentioned in this section the reference period is the baseline.

TABLE V
THE RESULTS OF THE MIXED LINEAR MODEL FOR BLINKS/M

	β	*t*	*p*
Period	- 7.64	87.00	< .001***
Biological sex	1.42	17.43	< .001***
Age	0.20	0.73	.40
Duration of the Recording	-0.02	0.03	.87
Age * Biological Sex	-0.32	3.06	.08

*** Refers to p-values < .001.

TABLE VI
THE RESULTS OF THE MIXED LINEAR MODEL FOR BLINKS' AVERAGE DURATION

	β	*t*	*p*
Period	-29.09	94.59	< .001***
Biological sex	1.31	0.12	.73
Age	-0.17	0.15	.70
Duration of the Recording	0.16	0.25	.62
Age * Biological Sex	0.11	0.05	.82

*** Refers to p-values < .001.

TABLE VII
THE RESULTS OF THE MIXED LINEAR MODEL FOR BLINKS' AVERAGE INTERVAL

	β	*t*	*p*
Period	4.96	19.15	.001***
Biological sex	1.13	7.86	.006**
Age	0.07	5.59	.02*
Duration of the Recording	-0.13	1.76	.18
Age * Biological Sex	0.10	0.45	.51

* Refers to p-values < .05 while ** refers to p-values < .01, and *** refers to p-values < .001

TABLE VIII
THE RESULTS OF THE MIXED LINEAR MODEL FOR AVERAGE RMSSD

	β	*t*	*p*
Period	11.11	22.11	< .001***
Biological sex	4.81	10.79	.002**
Age	0.28	10.35	.002**
Duration of the recording	-0.23	1.33	.25
Age * Biological Sex	0.16	0.26	.62

* Refers to p-values < .05 while ** refers to p-values < .01.

To summarize, our within subjects factor (period: baseline vs task) was found significant when looking at all the EAR-derived features collected during the task. In both blinks and average blinks duration there was a significant decrease during the task while for the average blinks' interval and the average RMSSD during the task there was a significant increase during the task (these results can be visualized in Fig. 3). For these two EAR-derived features we also found a positive effect of age where with the increase of age these two features also undergo an increase.

B.3 Baseline-corrected measures

Some of the confounders (such as age or biological sex) may affect the measures collected during the task's recordings (in our case The Sustainable Port gameplay). So. in this part of the analyses, we appraised if the baseline-corrected (obtained by subtracting the measures collected during the task from the baseline recording) measures are not affected by the variables we controlled for in the mixed linear model (age, biological sex, duration of the recording, and the interaction between biological sex and age). Our results show no effects of these predictors in the four chosen features (see Table IX and Table X). This suggests that baseline corrected features may be less influenced by confounders (such as age and biological sex) compared to the measures collected during the task.

TABLE IX
DESCRIPTIVE STATISTICS OF THE BASELINE-CORRECTED EAR-DERIVED FEATURES FOR MALES AND FEMALES

	Males	Females
Baseline-Corrected Blinks/m	M = 7.02 (SD = 7.06)	M = 8.26 (SD = 8.54)
Average Baseline-Corrected Blinks' Duration	M = 25.40 (SD = 29.01)	M = 32.87 (SD = 27.86)
Average Baseline-Corrected Blinks' Interval	M = - 6.27 (SD = 14.68)	M = - 3.62 (SD = 3.93)
Average Baseline-Corrected RMSSD	M = -13.78 (SD = 29.12)	M = -8.38 (SD = 12.51)

TABLE X
THE RESULTS OF THE ANOVA (F(1,88)) FOR THE FOUR BASELINE-CORRECTED FEATURES

	Biological Sex	Age	Duration of the Recording	Age * Biological Sex
Blinks/ m	*F*= 0.52 *p* = .47	*F* = 0.27 *p* = .60	*F* = 0.30 *p* = 0.58	*F* = 2.19 *p* = .14
Average Blinks' duration	*F* = 1.09 *p* = .30	*F* = 0.08 *p* = .78	*F* = 0.01 *p* = .93	F = 0.03 *p* = .86
Average Blinks' Interval	*F* = 1.08 p = .30	*F* = 0.60 *p* = .81	F = 0.17 *p* = .68	*F* = 0.41 *p* = .52
Average RMSSD	*F* = 0.83 *p* = .37	*F* = 0.11 *p* = .74	F= 0.13 *p* = .72	F = 0.16 *p* = .69

B.4 Discrimination using the EAR-derived features

After having evaluated the potential effects of confounders on the baseline-corrected features (baseline recording – task recording), we used a stratified 5-fold cross-validation with logistic regression to evaluate if the four defined features could be used to discriminate PoR employees from students. Table XI show the classification results for both the dummy classifier and the logistic regression model.

TABLE XI
THE RESULTS OF THE CLASSIFICATION TASK FOR THE DUMMY CLASSIFIER AND FOR THE LOGISTIC REGRESSION

	ROC AUC	PR AUC	F1
Logistic Regression	0.70 (SD = 0.11)	0.64 (SD = 0.17)	0.62 (SD = 0.05)
Dummy Classifier (baseline)	0.50 (SD = 0.00)	0.46 (SD = 0.14	0.46 (SD = 0.02)

After having investigated the classification metrics, the underlying effect of age was evaluated. First. we looked into the confusion matrix (see Table XII; the positive class refers to PoR employees while the negative one to students). Then we checked if the age of the participants misclassified was different than the age of the participant correctly classified. The analysis showed that, running a Welch t-test, there was no difference between PoR employees correctly classified and those misclassified ($t(24.46) = -0.54$, $p = .59$). Same results were found in the student group ($t(29.31) = -0.73$, $p = .45$).

TABLE XII
THE RESULTS OF THE CONFUSION MATRIX ACROSS THE 5-FOLD STRATIFIED CROSS VALIDATION

	Positive	Negative
Positive	16	12
Negative	14	41

To further investigate the use of the baseline-corrected features for this task, another logistic regression model was trained using the McFadden pseudo-R2 to evaluate the amount of variance explained by our features on the entire dataset (see Table XIII).

TABLE XIII
THE RESULTS OF THE LOGISTIC REGRESSION WITH VALUE P-VALUES AND COEFFICIENTS FOR ALL THE FEATURES

	B	SD	U	L	*t*	*p*
Blinks/m	-0.087	0.041	0.033	-0.167	-2.133	.03*
Baseline-corrected Average Blinks' Duration	-0.022	0.009	-0.015	-0.022	-0.253	.80
Baseline-corrected Average Blinks' Interval	0.165	0.087	0.335	-0.006	1.895	.06
Baseline-corrected Average RMSSD	-0.083	0.033	-0.019	-0.146	-2.528	.01*

* Refers to p-values < .05.

Overall, the model obtained a Pseudo-R2 of 0.11 which is statistically different than the intercept-only model ($p = .01$). More specifically, we found that baseline-corrected blinks and baseline-corrected RMSSD are the most significant features in discriminating the two groups where PoR employees had lower baseline-corrected blinks (M = 5.09, SD = 7.33) than students (M = 8.73, SD = 7.81) and overall bigger variations between one blink of another in terms of RMSSD (PoR employees: M = -12.38, SD = 23.36; students: M = -10.56, SD = 22.33). A Marginally significant difference ($p = .06$) was found in the average blinks' intervals (PoR employees: M= -4.01, SD = 8.59; students: M = -5.37, SD = 11.70) while no differences were detected in the average blinks' duration (PoR employees: M = 25.96, SD = 30.20; students: M = 30.44, SD = 27.91).

IV. Discussions

This study used EAR-derived features extracted to discriminate PoR employees from students playing a game simulating port dynamics. As in a previous study with The Sustainable Port, PoR employees performed better than students (after controlling for biological sex, age, video game habits, and digital game habits), suggesting that employees took advantage of their port-related knowledge to play the game [5]. Most PoR employees reported that the game simulates the dynamics of the Port of Rotterdam ecosystem and that playing the games raises awareness about the complexity of decisions in that ecosystem [5]. Our results, which relied on a multiple linear regression, suggest that the same patterns found in [5] can be also found when not considering the CO_2 threshold. This shows that, independently from reaching the threshold, PoR employees still have a significantly better performance, in terms of absolute score, when compared to students. Overall, given these findings, we could expect a difference in the two groups on a physiological level as well.

In this specific study, we used EAR-derived features extracted using a non-invasive method presented in a previous study [8], such as blinks/m, blinks' average duration, blinks' average interval, and RMSSD. Blinks/m have already been found to be significantly associated with performance on a task [16] and working memory [40], among other things. The general idea is that better-performing individuals, such as experts, blink more often per minute during a task when compared to their less experienced counterparts [8]. This may be due either to the connection that blinks/m have with working memory [40] or with the experienced workload [10]. A similar trend was found when looking at experienced video game players during a Hearthstone tournament [9], during a Tetris gameplay [8], and in some phases of a suturing task in a previous study involving surgeons [10]. With this being said, one of the problems of some of the studies mentioned is that they did not correct for the baseline at rest which, as seen in the current study, may play a role in determining the blinks/m during the task. We did that in ours.

Potential confounders when it comes to using the uncorrected blinks/m, or other variables as the ones we used in this study, are not limited to the baseline at rest but also to the age and the biological sex of the participants. As we saw, all of our features, besides the average blink duration, are affected by the biological sex of the participant. Furthermore, age proved to be a significant predictor of two of our features during the task (average interval duration and RMSSD); such differences may be due to age-connected variations in blinks' patterns found in other studies [30]. Besides the effect of age and biological sex we found that during the game session, there are significant variations in all the features we considered. More specifically, in our mixed linear models, we saw a decrease in the average blinks' duration, a decrease in blinks/m [8], and an increase in RMSSD, and average blinks' intervals [25]. The effect of playing video games, and overall facing a task that required attention, not only resulted in lowered blinks/m [8] but also in higher average blink intervals [25], higher RMSSD, and shorter blink duration [26]. Interestingly enough, when using the baseline-corrected measures we did not find any significant effect of the confounders we kept into account for this study. As can be seen in Table X, age and biological sex seem not to affect the baseline corrected features; this seems to be further confirmed when considering that no significant difference was found between players correctly classified and those missclassified. However, to further dissipate the possibility that age played a role in the identification of players future research should repeat this study collecting a more homogenous group of players (for what concerns age) where non-experts should have the same age as the PoR employees' group. Nevertheless, the results here obtained suggest that baseline-corrected measures are presumably more robust to the effect of confounders, and, consequently, we suggest their use in future studies. Furthermore, in this study, we did not find any difference in the baseline measures between the two groups under analysis (PoR employees and students controlling for age as covariate) suggesting the effects found are allegedly not due to baseline differences. These preliminary analyses about the effect of potential confounders, baseline-task variations, and baseline-corrected measures, were necessary before introducing a discussion about our results obtained using logistic regression models. Future studies should investigate the application of the method here proposed and evaluate their robustness when investigating other phenomena.

As we saw, our logistic regression, both on the stratified 5-fold cross-validation and the results of the Pseudo R2 show that it is possible to discriminate the two groups above baseline. More specifically, we found that baseline-corrected blinks and RMSSD are significant predictors in our logistic regression model. The results obtained in the RMSSD should be interpreted together with the average blinks' intervals which were found as marginally significant in our study. Taken together, these results suggest that PoR employee experience a global lower decrease in their average blinks' intervals but higher variations between one blink and another in terms of RMSSD. Such differences may be due to a better allocation of visual resources and better task-specific adaptability. Such interpretation follows the results obtained in heart rate variability studies where better performance in tasks is connected to higher RMSSD [41, 42]. For what concerns baseline-corrected blinks/m, our results are in line with what was obtained in a study focusing on performance-based expertise in Tetris [8]. A lower decrease in blinks/m may, for example, suggest either a lower experienced cognitive workload by the PoR employees [10] or better working memory in these specific tasks [40]. Another option is that given the higher performance showcased by PoR employees, variations in baseline-corrected blinks may be connected to the higher performance that this group had in the game as similarly found in other studies [8]. The results here obtained may provide evidence that PoR employees engage differently than students with The Sustainable Port given the possibility of using EAR-derived features and their connection with aspects such as working memory [40], and performance [16]. However, these hypotheses about the baseline-corrected RMSSD and the baseline-corrected blinks should be further investigated in future studies. As illustrated in this study, this is one of the first studies using baseline corrected blinks/m and other EAR-derived features. Future research may clarify the connection between these features and cognition.

To sum up, our results seem to suggest that it is possible to discriminate between experts in real life (PoR employees in our case) and laypeople (students) playing a game aiming to simulate a port environment. Such results provide evidence that a transfer between what is learned in real life and video games simulating real-life scenarios may occur. This is probably one of the most relevant assumptions when it comes to the use of serious games; their effectiveness implies that some real-life relevant behaviour may be transferred to games. Expertise has already extensively investigated with sensors such as electrocardiograms to track heart rate variability [11], electroencephalograms [12], or eye tracking [13]. This information while being effective in tracking expertise require often expensive sensors and knowledge to be understood and collected. The method here proposed requires widely available technology and it is based on openly available algorithms. Therefore, such a method provides an accessible tool for researchers with limited resources while at the same time proposing the base for the development of future business applications. However, the results here obtained concern only one game namely The Sustainable Port. Future studies may apply this method to other fields providing a more solid ground for the use of serious games for purposes such as recruitment or training in fields far beyond the mere application to port environments.

However, despite our results, several limitations affecting this study should be mentioned. First, our sample was relatively small and strongly imbalanced (this was also due to the natural distribution of experts in the population), and this may have affected our results. Second, we did not evaluate potential differences occurring between junior and senior employees (this is still due to a small number of PoR employees we managed to collect). Third, we did not control for other physiological variables that may play a role in discriminating PoR employees from students such as eye-tracker information. Information extractable such as saccades or fixations may provide a more exhaustive overview of what differentiates PoR employees from students as suggested in other studies using eye tracking to discriminate individuals with different levels of expertise in a task [43]. Adding such information may be also beneficial when performing a classification task possibly increasing scores obtained in the metrics to evaluate the performance of the classifier. Fourth, we focused just on age and biological sex as main potential confounders that may affect blinks and other EAR-derived features. This approach was used since the main point here was to provide a general robust methodology for blinks-related studies that can be applied to different fields and sectors. Future studies may focus on the effect of other variables such as skills related to working memory for example given the connection found in other studies. Future studies may repeat the current study with a bigger sample collecting more physiological measures.

The results obtained in this study not only provide insights into tracking expertise in serious games used for business purposes but constitute a first step towards applying this method in other fields. Previous studies show that methods similar to what employed in our study were used to track and detect expertise in video games [8]. Such a study obtained results similar to what we saw in the experiment presented here. The EAR-derived features used could be also employed to track expertise not only in (serious) games but also when it comes to other screen-presented tasks. Future studies, applying the methods presented here to study expertise, may obtain results similar to the ones obtained in our study showing that both expertise in real life and expertise (serious) video games present close physiological aspects. This hypothesis is based on the results of previous studies showing a decrease in blinks in expert surgeons during a specific phase of a suturing task [10]. As previously conveyed, one limitation of our study concerns the imbalance in classes, which is to be expected since experts represent a small portion of the population. However, the analysis pipeline adopted here shows that robust results can be still obtained despite class imbalance. Therefore, given the natural distribution of experts in the population, the analysis pipeline presented here could be successfully adopted in other studies aiming to investigate expertise. Consequently, the non-invasive methods suggested here can be used together with more classical methods, such as eye-tracking [13], to track and study expertise in other tasks beyond (serious) video games.

To summarize, in this study we found that expertise may be detectable above baseline using only baseline-corrected EAR-derived features which seem to be robust towards confounders such as age or biological sex. Future studies should focus on using other serious games to evaluate the reproducibility of our results. Furthermore, future research may also collect bigger samples than ours evaluating the effect of other potential influencing variables and combining other features extracted for example from eye-tracking to evaluate the effect this may have when training classifiers. Such research may be beneficial for detecting experts in real life, for example presented with a screen task, but also in detecting experts, or interesting profiles for hiring purposes, in simulations such as serious games.

V. Conclusions

The aim of this study was to investigate if physiological differences, in terms of EAR-derived features, can be used to discriminate in-real-life experts (PoR employees) from students. Our results suggest that EAR-derived features can be successfully used to discriminate the two groups. Future studies may extend the results of this work by collecting data with different serious games and including more physiological measures.

Acknowledgments

This research is founded by the MasterMinds project, part of the RegionDeal Mid- and West-Brabant, and it is co-founded by the Ministry of Economic Affairs and Municipality of Tilburg.

References

[1] H. R. Marston and M. del Carmen Miranda Duro, "Revisiting the twentieth century through the lens of Generation X and digital games: A scoping review," The Computer Games Journal, vol. 9, pp. 127-161, 2020.

[2] D. Muriel and G. Crawford, "Video games and agency in contemporary society," Games and Culture, vol. 15, no. 2, pp. 138-157, 2020.

[3] O. Allal-Chérif and M. Bidan, "Collaborative open training with serious games: Relations, culture, knowledge, innovation, and desire," Journal of Innovation & Knowledge, vol. 2, no. 1, pp. 31-38, 2017.

[4] E. Buiel, G. Visschedijk, L. H. E. M. Lebesque, I. M. P. J. Lucassen, B. Van Riessen, A. Van Rijn, and G. M. Te Brake, "Synchro mania-design and evaluation of a serious game creating a mind shift in transport planning," in 46th International Simulation and Gaming Association Conference, ISAGA, Jul. 2015, pp. 1-12.

[5] "Introducing The Sustainable Port: a serious game to study decision-making in port-related environments" (Currently in peer-review).

[6] A. Van der Gijp, C. J. Ravesloot, H. Jarodzka, M. F. Van der Schaaf, I. C. Van der Schaaf, J. P. van Schaik, and T. J. Ten Cate, "How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology," Advances in Health Sciences Education, vol. 22, pp. 765-787, 2017.

[7] M. Fenton-O'Creevy, J. T. Lins, S. Vohra, D. W. Richards, G. Davies, and K. Schaaff, "Emotion regulation and trader expertise: Heart rate variability on the trading floor," Journal of Neuroscience, Psychology, and Economics, vol. 5, no. 4, p. 227, 2012.

[8] G. Guglielmo, M. Klincewicz, E. H. in'Veld, and P. Spronck, "Tracking Early Differences in Tetris Performance using Eye Aspect Ratio Extracted Blinks," IEEE Transactions on Games, 2023

[9] G. Guglielmo, P. M. Blom, M. Klincewicz, E. M. J. H. in 't Veld, and P. Spronck, "Blink To Win," in Foundation of Digital Games 2022, 2022, doi: 10.1145/3555858.3555864 (former 18).

[10] R. Bednarik, J. Koskinen, H. Vrzakova, P. Bartczak, and A. P. Elomaa, "Blink-based estimation of suturing task workload and expertise in microsurgery," in Proceedings: 31st IEEE International Symposium on Computer-Based Medical Systems-CBMS 2018.

[11] M. Libertin, M. Ferguson, J. Gregory, and M. Zubrow, “702: Heart rate variability differentiates expertise level in point-of-care lung US interpretation,” *Critical Care Medicine*, vol. 52, no. 1, p. S323, 2024.

[12] P. Cantou, H. Platel, B. Desgranges, and M. Groussard, "How motor, cognitive and musical expertise shapes the brain: Focus on fMRI and EEG resting-state functional connectivity," *Journal of Chemical Neuroanatomy*, vol. 89, pp. 60-68, 2018.

[13] S.E. Fox and B. E. Faulkner-Jones, "Eye-tracking in the study of visual expertise: methodology and approaches in medicine," *Frontline Learning Research*, vol. 5, no. 3, pp. 29-40, 2017.

[14] P. P. Caffier, U. Erdmann, and P. Ullsperger, "Experimental evaluation of eye-blink parameters as a drowsiness measure," European journal of applied physiology, vol. 89, pp. 319-325, 2003.

[15] B. Zheng, X. Jiang, G. Tien, A. Meneghetti, O. N. M. Panton, and M. S. Atkins, "Workload assessment of surgeons: correlation between NASA TLX and blinks," Surgical endoscopy, vol. 26, pp. 2746-2750, 2012.

[16] R. Paprocki and A. Lenskiy, "What does eye-blink rate variability dynamics tell us about cognitive performance?," Frontiers in human neuroscience, vol. 11, p. 620, 2017.

[17] C. C. Liu, R. C. N. D’Arcy, T. Cheung, X. Song, and R. C. N. D’Arcy, "Spontaneous Blinks Activate the Precuneus: Characterizing Blink-Related Oscillations Using Magnetoencephalography.," Frontiers in Human Neuroscience, vol. 11, p. 489, Oct. 2017, doi: 10.3389/fnhum.2017.00489.

[18] Y. Wang, S. Toor, R. Gautam, and D. B. Henson, "Blink Frequency and Duration during Perimetry and Their Relationship to Test–Retest Threshold Variability," Investigative Ophthalmology & Visual Science, vol. 52, no. 7, p. 4546, Jun. 2011, doi: 10.1167/iovs.10-6553.

[19] A. Abusharha, "Changes in blink rate and ocular symptoms during different reading tasks," Clinical Optometry, vol. Volume 9, pp. 133–138, Nov. 2017, doi: 10.2147/opto.s142718.

[20] M. J. Doughty and T. Naase, "Further Analysis of the Human Spontaneous Eye Blink Rate by a Cluster Analysis-Based Approach to Categorize

Individuals With 'Normal' Versus 'Frequent' Eye Blink Activity," Eye & Contact Lens-science and Clinical Practice, vol. 32, no. 6, pp. 294–299, Dec. 2006, doi: 10.1097/01.icl.0000224359.32709.4d.

[21]R. Schleicher, N. Galley, S. Briest, and L. Galley, "Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired?," Ergonomics, vol. 51, no. 7, pp. 982–1010, Jun. 2008, doi: 10.1080/00140130701817062.

[22]S. M. Groman, A. S. James, E. Seu, S. Tran, T. A. Clark, S. N. Harpster, M. Crawford, J. L. Burtner, Karen Feiler, R. H. Roth, J. D. Elsworth, E. D. London and J. D. Jentsch, "In the blink of an eye: relating positive-feedback sensitivity to striatal dopamine D2-like receptors through blink rate," Journal of Neuroscience, vol. 34, no. 43, pp. 14443-14454, 2014.

[23]M. Brych, S. Murali, and B. Händel, "How the motor aspect of speaking influences the blink rate," PLOS ONE, vol. 16, no. 10, p. e0258322, Jan. 2021, doi: 10.1371/journal.pone.0258322.

[24]S. Leal and A. Vrij, "Blinking During and After Lying," Journal of Nonverbal Behavior, vol. 32, no. 4, pp. 187–194, Jul. 2008, doi: 10.1007/s10919-008-0051-0.

[25]A. Lenskiy and R. Paprocki, "Blink rate variability during resting and reading sessions," 2016, doi: 10.1109/norbert.2016.7547466.

[26]R. Mallick, D. Slayback, J. Touryan, A. J. Ries, and B. J. Lance, "The use of eye metrics to index cognitive workload in video games," in 2016 IEEE second workshop on eye tracking and visualization (etvis), Oct. 2016, pp. 60-64.

[27]L. Israel, P. Paukner, L. Schiestel, K. Diepold, and F. Schönbrodt, "Data for: Open Library for Affective Videos (OpenLAV)," PsychArchives, Aug. 12, 2021. https://www.psycharchives.org/en/item/009953b8-a55a4771-b25a-f6235bb159a2.

[28]S. Mehrang, E. Helander, M. Pavel, A. Chieh, and I. Korhonen, "Outlier detection in weight time series of connected scales," In 2015 IEEE international conference on bioinformatics and biomedicine (BIBM), (pp. 1489-1496). 2015, doi: 10.1109/bibm.2015.7359896.

[29]P. J. Rousseeuw and C. Croux, "Alternatives to the Median Absolute Deviation," Journal of the American Statistical Association, vol. 88, no. 424, p. 1273, Dec. 1993, doi: 10.2307/2291267.

[30]C. Sforza, M. Rango, D. Galante, N. Bresolin, and V. F. Ferrario, "Spontaneous blinking in healthy persons: an optoelectronic study of eyelid motion," Ophthalmic and Physiological Optics, vol. 28, no. 4, pp. 345-353, 2008.

[31]H. Schielzeth, N. J. Dingemanse, S. Nakagawa, D. F. Westneat, H. Allegue, C. Teplitsky, D. Réale, N. A. Dochtermann, L. Z. Garamszegi, and Y. G. Araya-Ajoy, "Robustness of linear mixed-effects models to violations of distributional assumptions," Methods in ecology and evolution, vol. 11, no. 9, pp. 1141-1152, 2020.

[32]A. Persson, H. Jonasson , I. Fredriksson , U. Wiklund , and C. Ahlström, "Heart rate variability for classification of alert versus sleep deprived drivers in real road driving conditions," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 6, pp. 3316-3325, 2020.

[33]J. S. Long and L. H. Ervin, "Using heteroscedasticity consistent standard errors in the linear regression model," The American Statistician, vol. 54, no. 3, pp. 217-224, 2000.

[34]P. Dudgeon, "Some improvements in confidence intervals for standardized regression coefficients," Psychometrika, vol. 82, pp. 928-951, 2017.

[35]A. F. Hayes and L. Cai, "Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation," Behavior research methods, vol. 39, pp. 709-722, 2007.

[36]L. A. Jeni, J. F. Cohn, and F. De La Torre, "Facing imbalanced data--recommendations for the use of performance metrics," in 2013 Humaine association conference on affective computing and intelligent interaction, Sep. 2013, pp. 245-251.

[37]H. Lu, L. Ehwerhemuepha, and C. Rakovski, "A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance," BMC medical research methodology, vol. 22, no. 1, p. 181, 2022.

[38]E. Richardson, R. Trevizani, J.A Greenbaum, H. Carter, M. Nielsen, and B. Peters, "The ROC-AUC Accurately Assesses Imbalanced Datasets." Available at SSRN 4655233, 2023.

[39]E. O. Bayman and F. Dexter, "Multicollinearity in logistic regression models," Anesthesia & Analgesia, vol. 133, no. 2, pp. 362-365, 2021.

[40]J. Ortega, C. R. Plaska, B. A. Gomes, and T. M. Ellmore, "Spontaneous eye blink rate during the working memory delay period predicts task accuracy," Frontiers in Psychology, vol. 13, p. 788231, 2022.

[41]K. Kaida, T. Åkerstedt, G. Kecklund, J. P. Nilsson, J. Axelsson, "Use of subjective and physiological indicators of sleepiness to predict performance during a vigilance task," Industrial health, vol. 45, no. 4, pp. 520-526, 2007.

[42]K. Hilgarter, K. Schmid-Zalaudek, R. Csanády-Leitner, M. Mörtl, A. Rössler, H.K. Lackner, "Phasic heart rate variability and the association with cognitive performance: A cross-sectional study in a healthy population setting," PLoS One, vol. 16, no. 3, p. e0246968, 2021.

[43]A. Gegenfurtner, E. Lehtinen, and R. Säljö, "Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains," Educational psychology review, vol. 23, pp. 523-552, 2011.