





Article

Correlating Personal Resourcefulness and Psychomotor Skills: An Analysis of Stress, Visual Attention and Technical Metrics

Carmen Guzmán-García ¹, Patricia Sánchez-González ^{1,2}, Juan A. Sánchez Margallo ³, Nicola Snoriguzzi ¹, José Castillo Rabazo ³, Francisco M. Sánchez Margallo ³, Enrique J. Gómez ^{1,2} and Ignacio Oropesa ^{1,*}

¹ Biomedical Engineering and Telemedicine Centre, ETSI Telecomunicación, Centre for Biomedical Technology, Universidad Politécnica de Madrid, 28040 Madrid, Spain; carmen.guzmang@upm.es (C.G.-G.); p.sanchez@upm.es (P.S.-G.); nicola.snoriguzzi@alumnos.upm.es (N.S.); enriquejavier.gomez@upm.es (E.J.G.)

² Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, 28029 Madrid, Spain

³ Centro de Cirugía de Mínima Invasión Jesús Usón, 10071 Cáceres, Spain; jasanchez@ccmijesususon.com (J.A.S.M.); jcastillo@ccmijesususon.com (J.C.R.); msanchez@ccmijesususon.com (F.M.S.M.)

* Correspondence: i.oropesa@upm.es; Tel.: +34-910-672-458

Abstract: Modern surgical education is focused on making use of the available technologies in order to train and assess surgical skill acquisition. Innovative technologies for the automatic, objective assessment of nontechnical skills are currently under research. The main aim of this study is to determine whether personal resourcefulness can be assessed by monitoring parameters that are related to stress and visual attention and whether there is a relation between these and psychomotor skills in surgical education. For this purpose, we implemented an application in order to monitor the electrocardiogram (ECG), galvanic skin response (GSR), gaze and performance of surgeons-in-training while performing a laparoscopic box-trainer task so as to obtain technical and personal resourcefulness' metrics. Eight surgeons (6 nonexperts and 2 experts) completed the experiment. A total of 22 metrics were calculated (7 technical and 15 related to personal resourcefulness) per subject. The average values of these metrics in the presence of stressors were compared with those in their absence and depending on the participants' expertise. The results show that both the mean normalized GSR signal and average surgical instrument's acceleration change significantly when stressors are present. Additionally, the GSR and acceleration were found to be correlated, which indicates that there is a relation between psychomotor skills and personal resourcefulness.

Keywords: stress; attention; personal resourcefulness; psychomotor skills; surgical skills; surgical education



Citation: Guzmán-García, C.; Sánchez-González, P.; Margallo, J.A.S.; Snoriguzzi, N.; Rabazo, J.C.; Margallo, F.M.S.; Gómez, E.J.; Oropesa, I. Correlating Personal Resourcefulness and Psychomotor Skills: An Analysis of Stress, Visual Attention and Technical Metrics. *Sensors* **2022**, *22*, 837. <https://doi.org/10.3390/s22030837>

Academic Editor: James F. Rusling

Received: 17 November 2021

Accepted: 21 January 2022

Published: 22 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Modern surgical education is focused on making use of the available technologies in order to train surgical residents, monitor their progress and assess their surgical skill acquisition [1]. Surgical skills are, nowadays, commonly split into either technical or nontechnical categories [2]. Technical skills comprise a series of competencies, such as psychomotor skills, swift eye-hand coordination and dexterity [3]. Nontechnical skills encompass varied competencies such as cognitive readiness [4], decision making [5], interpersonal abilities (i.e., communication and teamwork [6,7]) and personal resourcefulness (i.e., attention and stress management) [2,8]. This artificial bifurcation has helped to further categorize surgical competency into subcategories of skills.

The applicable technologies for technical skills' characterization (e.g., the instruments' or hand's motion tracking analysis) have been widely studied and effectively used in surgical educational environments [9–13]. Studies focusing on simulation have revealed that the objective assessment of psychomotor skills is feasible through the combination of motion-analysis-derived efficiency and quality metrics [14,15]. Efficiency metrics are

related to measurable physical parameters, which makes them objective and reproducible (e.g., time, path length, speed, etc.). These metrics can be divided into motion-derived metrics (e.g., acceleration, economy of area, economy of volume, etc.) and force-derived metrics (e.g., the force applied in sutures or tissue handling). Quality metrics relate to a task's definition and execution (e.g., number of errors, tissue damage, etc.).

However, as previously stated, surgical skills are not just technical in nature [16]. Nontechnical skills also play a major role when characterizing surgical skills; thus, it becomes necessary to develop automatic, objective monitoring and assessment methods for them. Innovative technologies are being developed, implemented and applied for the automatic, objective assessment of nontechnical skills [17]. For instance, visual attention and stress monitoring (especially based on wearable technologies) are gaining relevance within the surgical community due to their influences on personal resourcefulness [18].

Personal resourcefulness was initially defined by Brunckhorst et al. [19] as the ability to cope with stress and fatigue. Madani et al. [20] refined this definition by incorporating two main factors into it: (1) "self-awareness and metacognition" and (2) "management of modulators of attention, stress and goals". Essentially, personal resourcefulness encompasses the ability that a surgeon must have to "self-regulate one's cognitive processes and behaviours, with awareness of various modulating factors that can impact performance" [20]. Thus, personal resourcefulness includes maintaining focus for extensive periods of time amidst several distractions (e.g., sudden noises, surgical team movements, etc.), balancing attention between several intraoperative considerations (e.g., patient condition, performance of the surgical team, etc.) and managing the modulators of stress or attention [20,21].

1.1. Modulators of Attention

Visual attention and hand-eye coordination skills are essential for developing and mastering surgical techniques [22–24]. Different experimental studies have demonstrated a close relationship between the eye movements that are occurring during visual attention tasks and the way in which the information is processed at a cognitive level [25,26]. Some studies have shed light on the visual attention patterns between experts and nonexperts (based on the classification proposed by Land et al. to differentiate between experts and nonexperts [27], experts have more than 5 years of experience, while nonexperts have less than 5 years of experience), showing that, while experts tend to fix their attention on the target, nonexperts show rather erratic behavior, often focusing attention on their own instrument [22–24]. These studies suggest that a more efficient attention ability (i.e., the ability to shift the attention's focus to areas requiring higher consideration) could give ground to other cognitive processes during surgical interventions. Given the importance of visual attention, it becomes essential to develop learning strategies to train and assess the skills that are related to visual attention and concentration. For instance, Wilson et al. [24] designed and validated the effectiveness of a surgical performance enhancement program that is based on gaze training by means of a measure of target locking (computed by subtracting the percentage of time spent fixating on the surgical tool from the time spent fixating on the target). Additionally, Richstone et al. [28] showed that eye metrics, such as pupil diameter, blink rates and fixation rates, could effectively differentiate between surgical experts and nonexperts (i.e., using linear discriminant analysis and neural network analysis) and, thus, might be used as an objective measurement of surgical skill.

Visual attention can also be used to train other surgical staff. Koh et al. [29] studied the differences in eye movements between expert and nonexpert nurses in cesarian section surgeries. Their study found that the expert nurses distributed their attention more effectively during high-stress stages of the surgery by better focusing on the important areas of interest for each stage. Areas of interest were defined by the distinct boundaries within which the scrub nurses perform their tasks depending on the stage of the surgery. Experts also exhibited increased performance (measured by means of Objective Structured Assessment of Technical Skills—OSATS—questionnaire). In a different experiment, Tomizawa et al. [30] examined the eye movements of perfusionists, perceiving that those with less experience

visually inspected key information areas less often than the more experienced ones (who also distributed their attention more widely, exhibiting lower fixation rates). Additionally, gaze training has been demonstrated to enhance the acquisition of laparoscopic technical skills [24].

We hypothesize that attention-based training's benefits can be increased by means of objective assessment methodologies. Currently, eye-tracking devices [31,32] and camera systems [33–35] (among other methods) are being used to monitor the surgeons' visual attention, as well as their fatigue and concentration levels. However, none are being applied in order to bring additional information about the surgeons' personal resourcefulness.

1.2. Modulators of Stress

There is evidence proving that intraoperative stress can affect the overall performance of surgeons (e.g., by complicating the communication and hampering the hand-eye coordination, as well as other motor functions) [36]. This may lead to inferior patient outcomes and bring about potential intra- and postsurgical complications [37]. Thus, the training of stress management skills can be said to have especial relevance, as surgeons will be more prepared to effectively manage stressful situations when encountering them.

Simulation-based training methods for stress management are used in the literature within laparoscopic training programs [38,39], specifically in box trainer- or mannequin-based tasks. Repeated simulation training in high fidelity settings [40] and the training of eye gaze under high-anxiety conditions [41] have also been used to train stress management. In addition, studies of mental training methods have demonstrated positive effects on participants' stress experience and a reduction of their cognitive stress [42–45], as well as an improvement of their technical performance [39,44,46,47]. Finally, in the study by Lemaire et al. [48], a biofeedback-based stress management tool was designed and its effectiveness was demonstrated through a randomized control trial in which the mean stress score declined significantly for the intervention group. This tool monitors certain stress-related parameters (i.e., heart rate and blood pressure) and alerts surgeons when surpassing threshold levels so that they can apply their stress management techniques.

Based on the usefulness of this biofeedback-based stress management tool, numerous studies in the literature have focused on the characterization of stress in the surgical practice by means of simulation tasks (e.g., virtual reality or box trainers) including stressful situations (such as unexpected bleeding or sudden noises), so that residents are able to learn how to deal with such situations.

Due to the simplicity of its measurement, heart rate (HR) has been used in several studies via an electrocardiogram (ECG)-based stress-monitoring method [37,49–51]. HR has been demonstrated to increase when performing a surgery and even more so in laparoscopic surgery (as compared to open interventions) [50]. Some studies have also demonstrated the potential of HR to differentiate between expert and nonexpert surgeons [36]. In a study by Marrelli et al. [52], HR was found to be higher in nonexpert surgeons than in experts. They suggested that this might be an indicator of the expert surgeons' enhanced stress management skills. Even though they recognize that the difficulty of the surgical procedure can influence cardiovascular alterations and cortisol levels, the way in which they are affected is significantly higher in junior surgeons than in senior surgeons.

Other ECG-derived metrics are related with HR variability (HRV), which is defined as the variation in the interval of time between the heart beats. Weenk et al. [53] and Grantcharov et al. [54], in their respective studies, evaluated how the standard deviation of the interval between two heart beats (SDNN) and the square root of the mean interval between two heart beats (RMSSD) were affected by stress levels during surgical performance. Both concluded that these two metrics decreased significantly during laparoscopic surgical procedures.

Typical HRV-derived metrics in the frequency domain that are used in surgical learning environments are related to the power in two bands: low frequency (LF) (range 0.05–0.15 Hz [55,56]) and high frequency (HF) (range 0.16–0.45 Hz [56]). This distinc-

tion is of particular importance since LF is commonly associated with the activity of the sympathetic nervous system (which triggers the stress response) [57]. Thus, the most popular frequency-based metric is the ratio between the absolute power of the signal in the low and high frequency bands (LF/HF). This ratio has been demonstrated to increase significantly when presented with stressful situations (such as surgical procedures) [50,53].

Despite the close relation observed between HR and HRV with stress, this monitoring method can be influenced by other factors since it measures both the stress response and HR fluctuations of the individual that are not necessarily related to stress (e.g., caffeine intake, arrhythmias, etc.). This is why additional monitoring methods are usually coupled with the analysis of ECG-derived metrics.

For instance, measurement of the electrodermal activity or the use of the galvanic skin response (GSR) have been reported in surgical educational environments for stress monitoring. The potential obstruction to the normal performance of the task when using GSR electrodes (usually placed in the fingertips) can be considered a major drawback. However, new electrodes have been developed and are being improved in order to measure the GSR via a wrist-band [58].

As an example of the use of the GSR in surgical environments, Smith et al. [59] proposed a workstation to monitor surgeons' mental workload that is associated with surgical procedures (i.e., the amount of mental resources that are required to perform a set of concurrent tasks [60]) with their GSR as an indicator of their mental stress levels while performing a suturing task. The data that were obtained for this parameter were correlated to self-rated levels of stress during the task, resulting in an accurate characterization of stress levels in surgical contexts (it was found that GSR increases when stress does). In the experiment of Ershad et al. [61], experts showed lower variability in the GSR signal, while nonexperts showed higher variability.

The metrics that are related to stress that can be extracted from GSR measurements that are found in the literature are varied in their complexity. The most popular approaches are mainly descriptive (i.e., the mean, standard deviation and sum of the GSR signal). For instance, Hurley et al. [62] used these metrics to find the differences in the stress levels between standard laparoscopic surgery and robotic surgery. These metrics have also been used to validate their relation with other variables, specifically with workload [63] and performance [64].

A different approach consists of analysing the peaks of the GSR signal. An example of this is the study by Wilson et al. [65] in which they discovered that the mean peak values are correlated with the difficulty of the task and that the nonexpert fellows presented mean peak values that were higher than those of the expert fellows.

A more complex processing of the GSR signal was proposed by Bakker et al. [66], who used a solution consisting of an aggregation step in the time axis and a discretization step for the GSR signal. The aggregation in time was performed by using the overlapping time windows to segment the signal and extract the mean values for each segment. With the discretization, only a finite number of the GSR values were considered. This approach has been found to be more robust and reliable in the presence of noise and artifacts; but, being that the processing is more complex, it is not much used in the literature [67].

Besides ECG and GSR, other monitoring systems have been used in order to monitor stress in a surgical educational context, though their use is less frequent, mainly due to their higher intrusiveness. For instance, in the study by Bartolomeo et al. [68], both HR and surface electromyography (EMG) were simultaneously measured by electrodes that were located on the left and right upper trapezium. They proved that muscle activation is much higher in live procedures than in virtual environments.

Electroencephalography (EEG) has seldom been used as a stress monitoring method in the surgical educational context, mostly due to the intrusive nature of EEG and the difficulty in the stress characterization in a specific frequency band. An example of the metrics that can be extracted from EEG in order to monitor stress in surgical environments is the ratio of the beta band over the alpha band. Duru et al. [69] designed an experiment in

order to monitor the mental workload during a simple laparoscopic nephrectomy using this metric, effectively characterizing the most difficult and stressful phases in the intervention. Meanwhile, Morales et al. proved that higher stress levels and surgical complexity are associated with higher activity in the beta-frequency of the EEG [70].

Lastly, functional near-infrared spectroscopy (fNIRS) has gained acceptance as a stress-monitoring method due to its applications in cognitive and behavioral studies [71,72]. fNIRS is an indicator of the activity of the prefrontal cortex (PFC), which is activated (among other cases) when presented with stressors. Specifically in surgical training environments, Shetty et al. demonstrated that nonexperts had higher levels of PFC activity than experts [73].

Other studies have focused on the assessment of stress during surgical procedures based on standard scales, such as the State–Trait Anxiety Inventory (STAI) [74–77] or ad hoc scales for surgical interventions [38,78]. Some studies have used the STAI as a correlation method for comparison against the physiological indicators of stress [38,75–77].

For instance, Arora et al. [75,79] designed a stress assessment tool (the imperial stress assessment tool—ISAT), by which quantitative data (HR and cortisol levels) and qualitative data (stress self-assessment by means of the STAI test) were gathered. ISAT was validated by scheduling 54 procedures varied in complexity (and thus, the level of stress that was induced), with the same 11 surgeons. The results of their validation indicated that both STAI and HR were higher for stressful procedures as compared to non-stressful ones. This suggests that both are highly correlated to the stress response. In another study, they added technical performance measurements [74] by using a MIST-VR simulator [80] by which the path length, total simulation time and number of errors could be derived. Both STAI and HR correlated with the performance measurements.

1.3. Goals of This Study

As observed in the literature, there are numerous studies that have focused on the characterization of stress and visual attention in surgical educational environments. Furthermore, it has been demonstrated that the application of monitoring technologies characterizing stress can be used to train personal resourcefulness (i.e., to recognize and manage modulators of stress and attention).

The tasks that are used for personal resourcefulness training are greatly varied and depend on the specific learning objective as well as the goals of the study itself. For instance, Platte et al. [81] used the valid and reliable star-track test [82] mounted on a box trainer to measure surgical performance while presented with the Norinder arithmetic test (as a stress-triggering method) in order to look for differences in performance between stressful and non-stressful situations in terms of the number of errors and the amount of time that was spent on the task. In the Norinder test, participants are required to add and subtract a series of paired one or two digit numbers under a certain time threshold. The Norinder arithmetic test has been demonstrated to trigger a stress response in subjects [83,84]. The results indicated that there were indeed differences. However, the individuals' stress was not quantitatively assessed (i.e., the physiological parameters that are correlated to the stress response were not monitored), which does not allow effective demonstration that the stress response was triggered during the task.

In addition, few studies have explored the potential of monitoring parameters for the objective and quantitative assessment of personal resourcefulness or whether these two skills are related to one another (i.e., whether the stress- and attention-based parameters may be related with other surgical skills metrics, such as the use of motion-analysis-based parameters to characterize psychomotor skills).

Thus, this present study has two main goals: (1) to quantitatively characterize the stress response during a basic surgical training task, and (2) to find whether the metrics that are related to psychomotor skills and personal resourcefulness in surgical education are correlated.

To verify or reject these hypotheses, we present an analysis of metrics that were derived from physiological parameters related to stress, gaze and laparoscopic instrument tracking in order to study how they are influenced by stressors and how they influence one another.

2. Materials and Methods

2.1. Subjects

Laparoscopic surgeons collaborating with the Centro de Cirugía de Mínima Invasión Jesús Usón (Cáceres, Spain) participated in the study after providing informed consent. Demographic information was gathered from each participant (i.e., age, gender and experience level). The participants were grouped, according to their years of previous surgical training experience, into experts (>5 years of experience) and nonexperts (<5 years of experience) [27].

The participants did not take part in any vigorous activity within 30 min of the task performance in order to avoid a bias in the ECG measurements. The exclusion criteria included concurrent use of medications affecting heart rate.

2.2. Experimental Setup

In order to allow the participants to engage their personal resources (i.e., to recognize and manage modulators of stress and attention), a specific task that was based on the star-track test [82] (see Figure 1) and the Norinder test was carried out. This combination was selected in light of the results that were obtained by Platte et al. [81] in an attempt to confirm their results via the quantitative assessment of stress. The addition of the Norinder test, in which stress and divided attention are triggered, assesses personal resourcefulness (since it includes the management of modulators of stress and attention). Furthermore, the cognitive demands that are necessary to solve the Norinder test can be associated with those of real surgical environments (only to a lesser extent) [85].

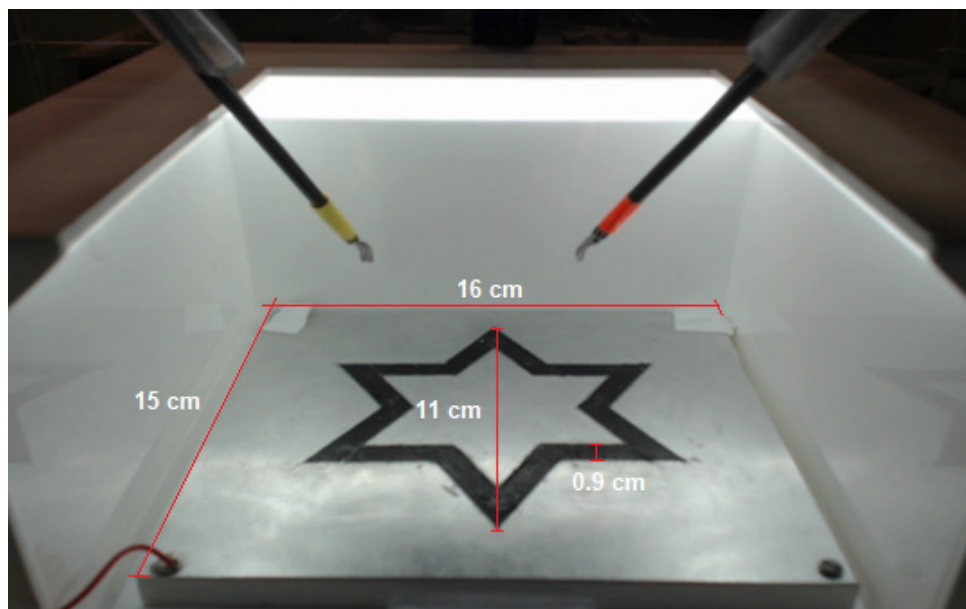


Figure 1. Star-track test (as seen in the box trainer).

The star-track test consists of following a star pattern with the laparoscopic instrument within a box trainer (using the dominant hand), as efficiently and quickly as possible and within the star's borders [81]. More specifically, the star pattern was created using an aluminum plaque (16 × 15 cm), on top of which a 0.9 cm-wide strip of duct tape was placed, forming a star pattern (point to point distance: 11 cm).

None of the participants were familiarized with the task nor allowed to practice it beforehand. For the first phase of the experiment, the participants were asked to complete

the star-track test and reminded to keep the instrument in continuous contact with the plaque. This had to be done a total of 10 times: 5 times counterclockwise and 5 times clockwise, as defined in the study by Platte et al. [81].

The second phase of the test was identical to the first but it included the Norinder arithmetic test and was carried out immediately after the first phase. The participants were encouraged to do the Norinder test with as few errors as possible. The test was always the same for all of the participants. The mathematical questions of the Norinder test had to be solved in under 5 s. When answered (or if not answered in time) the next question was shown. The length of the test (i.e., the number of additions and/or subtractions that were presented) depended on the time that the participant took to complete the 10 laps of the star-track test. The Norinder test was presented on a different screen from the main one on which the video feed from the inside of the box trainer was shown and each question was read out loud by the research assistant. This slightly diverted the attention of the surgeons while triggering the stress response. The results of the Norinder test (as mentally calculated by the participants) were verbally stated by the participants and manually introduced (via keyboard) by the research assistant. They were automatically saved in a text file and then compared with the actual answer in order to avoid participants randomly answering them. Random answers would indicate that no stress was triggered because no actual mental calculations were taking place and these results would be discarded from the analysis. In order to evaluate for randomness, the participants' answers were compared with the correct answers. If few questions were answered correctly (less than 15%), it was concluded that the answers were given at random and, as a result of this, the data were discarded.

Additionally, the level of stress that was perceived by the participants was collected on a Likert scale from 1 to 5 (1 being not stressed at all and 5 being very stressed) both before the experiment and after its completion.

2.3. Monitoring Application Description

For the experiment, a combination of the (1) physiological signals related to stress (i.e., ECG and GSR), (2) gaze coordinates of the participants on the screen that were related to visual attention and (3) motion analysis parameters (MAPs) that were obtained from the tracking of the dominant laparoscopic instrument was obtained (Figure 2).

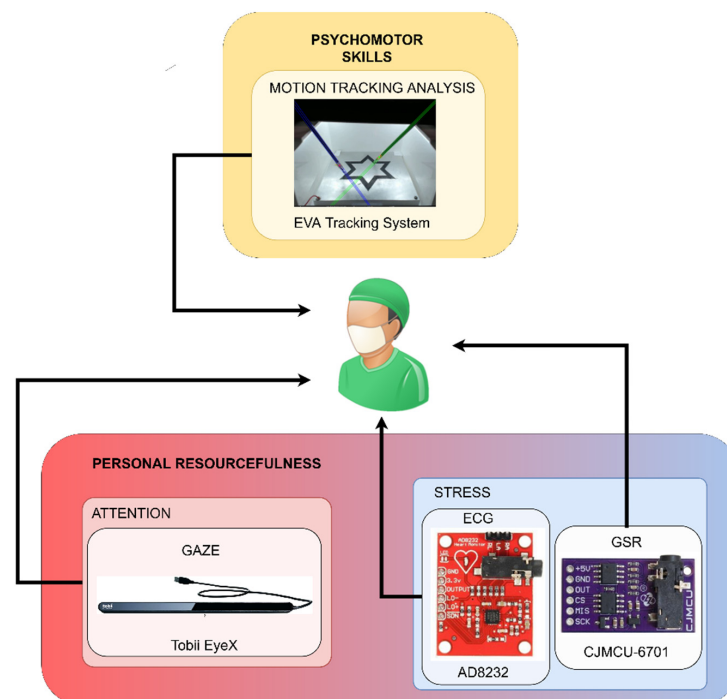


Figure 2. Monitoring parameters in the designed experiment.

2.3.1. Stress Module

To monitor stress, we used ECG and GSR commercial measurement devices in order to measure the HR and sweat gland activity, respectively. These two physiological signals were selected due to their growing acceptance and accuracy [36,86] as well as their demonstrated reliability in preliminary tests in surgical settings. The incorporation of EMG to measure muscular activity in the trapezius was originally considered. Contrary to the ECG and GSR, the EMG was discarded after the preliminary tests revealed no response towards the math problem-based stressor. Additionally, both the ECG and GSR were not intrusive for surgeons (unlike an EEG, for instance). More to the point, the electrodes were located in order to ensure that the cables did not interfere with the normal performance of the surgeon (Figure 3).



Figure 3. Setup of the experiment. ECG and GSR sensors are located inside the box trainer (red light). Eye tracker located immediately below the monitor showing the video feed captured by the camera inside the box trainer. The laptop is running the application gathering data. Research assistant is ready to start the second phase of the experiment, in which he will read mathematical questions out loud.

The devices were integrated on a Wemos D1 mini microprocessor with a sampling rate of 100 Hz. The total sample size depended on the time that was taken to complete the experiment, with an average of 289,000 samples having been yielded per subject. The sensor that was used to monitor the ECG was the AD8232 (SparkFun, Niwot, CO, USA). It consists of an integrated signal conditioning block for ECG [87] that is designed to extract, amplify and filter the small ECG signal in the presence of noisy conditions, such as muscle motion. To monitor the GSR, the CJMCU-6701 (DIY Electronics, Durban, KZN, SA, South Africa) [88], a signal conditioning block for GSR, was used. Both of the sensors were connected to different pins of the Wemos D1 mini microprocessor.

To measure the ECG, snap wet gel electrodes were used since the amount of the electrode's surface that is in contact with the skin is higher, thus providing a more accurate reading. The ECG electrodes were located under the right and left pectoral muscles and immediately below the rib cage [89]. These locations were selected due to the reduced amount of movement artifacts that are caused by muscular activity. The electrodes that were used to measure the GSR were located in the proximal phalanx of the index and middle fingers in order to achieve valid values.

The raw signal that was obtained by the ECG sensor was processed in order to obtain metrics related to the HRV and to the frequency-domain. More concretely, the signal was first filtered by means of a Butterworth high-pass filter (order = 3, $F_c = 20$ Hz) in order to remove the noise or distortion that was present in the signal. Baseline variations (i.e., respiratory noise, mean value due to electrodes contact, etc.) were then removed by means of an asymmetric least squares smoothing. This method uses a smoother with an asymmetric weighting of the deviations in order to get a baseline estimator. In doing so, the baseline can be removed while retaining the signal peak information. The resulting filtered signal was standardized (i.e., the values of the filtered signal were subtracted by the mean of the complete signal and divided by its standard deviation). After this, the local maxima were found (i.e., the peaks corresponding to the QRS complexes) by comparing neighboring amplitudes.

Once the QRS complexes were found, the NN intervals were computed by finding the distance between the peaks. The HRV signal was derived from the NN intervals and the time between them. The fast Fourier transform was applied to the interpolated HRV signal in order to select the high and low frequency components. The metrics that were used in this study were calculated according to the formulas that are shown in Table 1.

Table 1. Metrics of stress derived from ECG and HRV. Where t_{NNi} denotes the time intervals between two consecutive R peaks, x_{LF} is the filtered version of the HRV signal in the frequency band [0.05–0.15] Hz, T is the period of the signal and x_{HF} is the filtered version of the HRV signal in the frequency band [0.16–0.50] Hz.

Metric	Description	Formula
Time-domain metrics		
AVNN	Average value of NN intervals (ms).	$AV_{NN} = \frac{1}{N} \sum_{i=1}^N t_{NNi}$
SDNN	Standard deviation of NN intervals.	$SD_{NN} = \sqrt{\frac{1}{N} \sum_{i=1}^N (t_{NNi} - AV_{NN})^2}$
RMSSD	Root mean square of successive differences between successive NN intervals.	$RMS_{SD} = \sqrt{\frac{1}{N} \sum_{i=1}^N t_{NNi}^2}$
PNN20	Ratio of the number of pairs of consecutive NN intervals differing by more than 20 ms over the total number of NN intervals.	$P_{NN20} = \frac{ \{i \mid t_{NNi} > 20ms\} }{total\ No.\ of\ NN\ intervals}$
PNN50	Ratio of the number of pairs of consecutive NN intervals differing by more than 50 ms over the total number of NN intervals.	$P_{NN50} = \frac{ \{i \mid t_{NNi} > 50ms\} }{total\ No.\ of\ NN\ intervals}$
Frequency-domain metrics		
LF	Absolute power of the signal in low frequency bands.	$LF = \frac{1}{T} \sum_{t=1}^T x_{LF}^2(t)$
HF	Absolute power of the signal in high frequency bands.	$HF = \frac{1}{T} \sum_{t=1}^T x_{HF}^2(t)$
LF-HF ratio	Ratio between the total energy in low frequency and the total energy in the high frequency.	$LF - HF\ ratio = \frac{LF}{HF}$
LFnu	Normalized spectral LF index.	$LF_{nu} = \frac{LF}{LF + HF}$
HFnu	Normalized spectral HF index	$HF_{nu} = \frac{HF}{LF + HF}$

The raw signal that was obtained by the GSR sensor was normalized in order for it to be compared between the participants. Only the mean values were calculated as a metric of the GSR, since other validated metrics were not relevant for the specific data of the proposed experiment. More specifically, the slope was discarded after analysing each individual GSR signal and finding no common pattern in the slope, while peak analysis seemed irrelevant since the stress was continuous along the experiment instead of having been inflicted at given instants.

2.3.2. Attention Module

To monitor attention, we used the commercial eye-tracker Tobii EyeX (Tobii AB, Stockholm, Sweden). Tobii EyeX is an eye-tracking device that uses infrared light to track the pupil's motion, even when wearing glasses. It is placed at the bottom of the screen and can track the eye movement, the eye position and the gaze (i.e., the point in the screen that the person is looking at). In our particular case, we only used the gaze coordinates in order to compute the metrics to be correlated to personal resourcefulness. Before starting the experiment, the eye tracker was calibrated for each participant. To do this, they were asked to follow, with their gaze, a series of dots that were shown on the screen. The calibration tool (which is incorporated within Tobii's software, version 2.16.5) automatically fitted the parameters for better recognition of the participant's pupils.

The visual attention metrics that were extracted from the eye-tracker were the fixation rate (i.e., the rate at which at least two gaze points were at a maximum distance of 30 pixels apart), the average fixation duration (i.e., the amount of time a fixation was given) and the gaze ratio inside a region of interest (ROI)—namely the star pattern that was to be followed (Figure 4).



Figure 4. Example of the ROI that was used to extract the ratio of gaze points.

2.3.3. Performance Module

To monitor performance we used the Endoscopic Video Analysis (EVA) tracking system [12]. The EVA system enables the recording of the laparoscopic instruments' motion based on a nonobtrusive video tracking algorithm. Specifically, the EVA system employs information about the laparoscopic instruments' shaft edges in the image, their insertion point and the camera's optical center in order to track the 3D position of the instrument's tip. Thus, the system allows the analysis of the instruments' movements and renders a set of MAPs for the effective assessment of surgical psychomotor skills using the information that is obtained from a single camera. These MAPs (Table 2) were used to analyze the correlation of psychomotor skills with personal resourcefulness.

Table 2. MAPs obtained from EVA tracking system.

MAPs.	Definition	Formulae
Time (T)	Total time to perform a task (s)	T
Path length (PL)	Total path covered by the instrument in the setting (m)	$PL = \int_{t=0}^T \frac{d r(t) }{dt} dt$
Average speed (S)	Rate of change of the instrument's position in the setting (mm/s). Results are measured for the total magnitude and in each Cartesian direction of the box trainer.	$S_{u(t)} = \frac{1}{T} \int_{t=0}^T \frac{d u(t) }{dt} dt$
Average acceleration (A)	Rate of change of the instrument's velocity within the setting (mm/s ²)	$A = \frac{1}{T} \int_{t=0}^T \frac{d^2 r(t) }{dt^2} dt$
Economy of area (EOA)	Relationship between maximum surface area (task plane) occupied by the instrument and total path length	$EOA = \frac{\sqrt{SU^*}}{PL}$
Economy of volume (EOV)	Relationship between maximum volume occupied by the instrument in the setting and total path length	$EOA = \frac{\sqrt[3]{MV^{**}}}{PL}$
Depth (D)	Total path length travelled in the instrument's axis direction (m)	$D = \int_{t=0}^T \sqrt{\frac{dy(t)}{dt}^2 + \frac{dz(t)}{dt}^2} dt$

$$* SU = [max_t(x(t)) - min_t(x(t))] \cdot [max_t(z(t)) - min_t(z(t))] \quad ** MV = SU \cdot [max_t(y(t)) - min_t(y(t))].$$

Calibration of the EVA Tracking System had to be done beforehand using the video that was recorded during the experiment. This calibration essentially consisted of selecting the appropriate values in order to ensure that the EVA algorithm only recognized the marker of each laparoscopic instrument. This was manually done for each of the videos that were recorded during the experiment. Once the calibration was achieved, the corresponding video was fed to the EVA system for its coordinate and MAPs extraction. The MAPs were automatically calculated and uploaded to MongoDB (MongoDB Inc., New York, NY, USA) [90].

2.4. Data Analysis

To test for normality, the Shapiro–Wilk test was used for all of the metrics. In addition, QQ plots for all of the metrics were analyzed in order to ensure that the distribution was indeed normal or not normal (to account for the low sample size). Descriptive statistics for each of the participants are presented as a mean with standard deviation (SD) or a median with range, depending on the normality of the data's distribution. The values of the MAPs, gaze analysis and physiological metrics of the first phase of the experiment were compared with those of the second phase using the ANOVA or Mann–Whitney test, depending on the normality of the data. These were done in order to analyze whether there were significant differences in the presence of stressors during the laparoscopic practice; thus, to know if the analyzed metrics were a good indicator of stress.

Additionally, a correlation analysis was performed, according to either Pearson's or Spearman's correlation coefficient, depending on the normality of the data, in order to find correlations between the psychomotor skills and personal resourcefulness skills metrics.

We adopted a critical α level of 0.05 to keep false discoveries at a traditional level as we sought to improve the reliability and effect size of the tests for the hypothesis. We used the Benjamini–Hochberg [91] method in order to provide a false detection error rate correction.

All of the statistical tests were carried out using RStudio 1.3.1093 (RStudio, Inc., Boston, MA, USA).

Lastly, we conducted a post hoc power analysis using G*Power [92,93] in order to find out whether the experiment had enough power to (1) detect the differences in the parameters under the stressors' effects and (2) correlate the parameters of personal resourcefulness and psychomotor skills.

3. Results

In the initial experimental design, 20 surgeons were to be recruited. However, due to COVID-19, these numbers were reduced, mainly due to the unavailability of the surgeons. Thus, we recruited a total of 13 surgeons. However, due to unforeseen technical issues of

the application recording the data (i.e., USB disconnection mid-task), some of the data from the stressful phase from 5 surgeons were not correctly recorded and were not included in the analysis. Therefore, eight surgeons were included in this study, (4 males and 4 females, aged 32 ± 5), following the protocol that was described in Section 2.1.

Two of the participants were considered experts (>5 years of experience). The remaining six participants were nonexperts. The mean perceived stress levels (\pm SD) before their participation in the experiment (on a 5-point Likert scale) were $1.423 (\pm 0.847)$ whereas, by the end of the experiment, they were $3.975 (\pm 0.814)$, increasing by 179% ($p < 0.001$). The analysis of the answers to the mathematical questions revealed no random patterns; thus, the data from all of the participants were included in the study.

The average speed of acceleration significantly increased by 15.09% in the stressful phase of the experiment as compared to the first phase ($p < 0.05$) and the GSR increased by 261% ($p < 0.005$) (Figure 5). The post hoc power analysis indicated that the effect size that was obtained for the average acceleration was 1.59 and 2.97 for the GSR (these are large effects, according to Cohen's effect size conventions [94]). The powers that are required to detect effects of this size were determined to be 0.819 and 0.999, respectively. The remaining metrics did not present significant differences between the experimental phases. For further information, please see Table S1 in the Supplementary Materials.

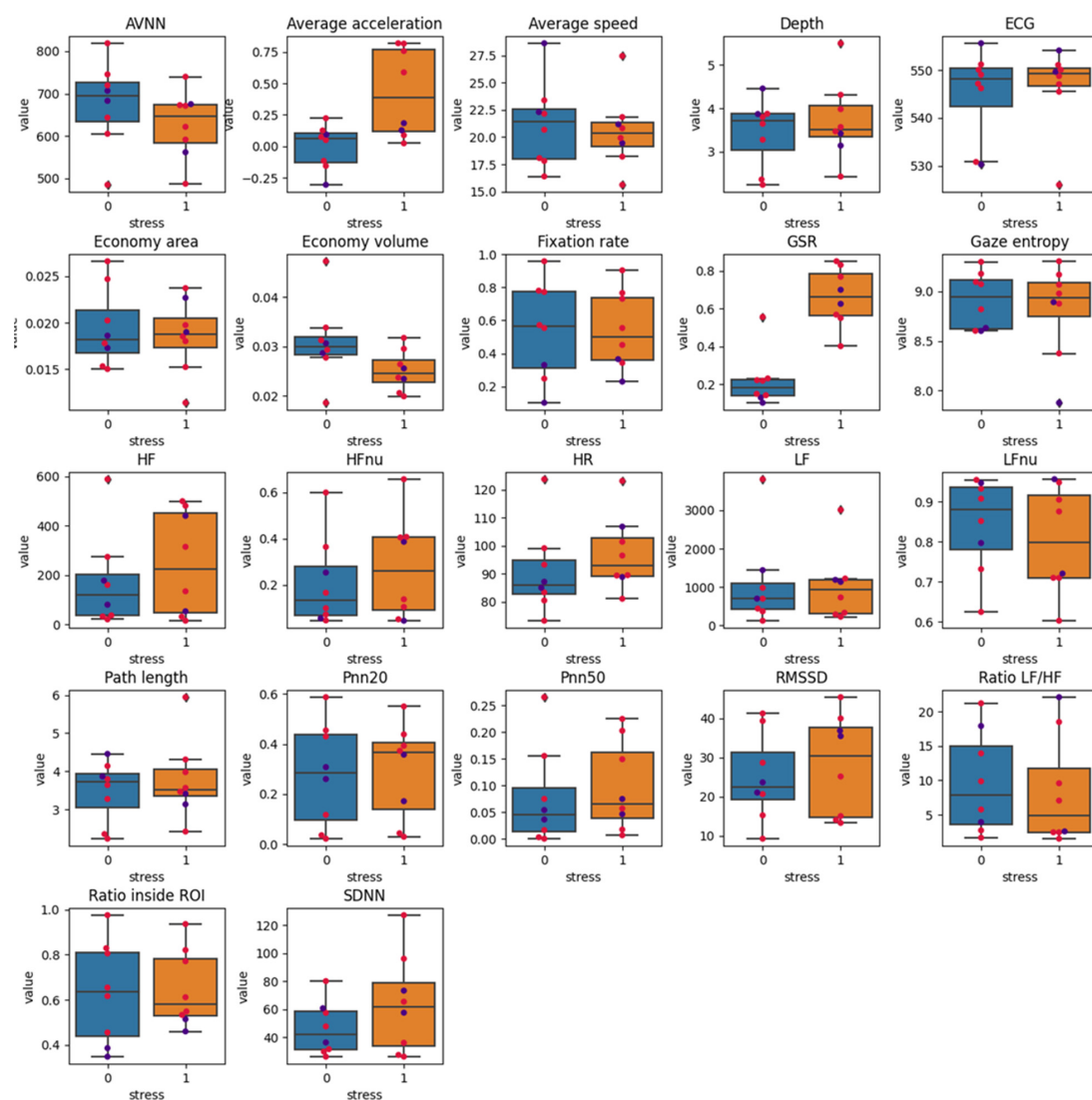


Figure 5. Box plot of phases 1 and 2 of the experiment for each of the metrics. Red and blue dots indicate the value for non-experts and experts, respectively.

The correlation analysis suggests that the GSR scores are correlated with the subjective stress that was stated by the participants at the beginning and end of the experiment ($\rho = 0.79, p = 9.99 \times 10^{-4}$). The GSR was also shown to correlate with the average acceleration ($\rho = 0.82, p = 2.017 \times 10^{-6}$). The post-hoc power analysis showed that the power for the correlation between the GSR and the subjective stress level is 0.99 (effect size = 0.88); and between the GSR and acceleration is 1 (effect size = 0.9).

The complete correlation results are reported in Figure 6.

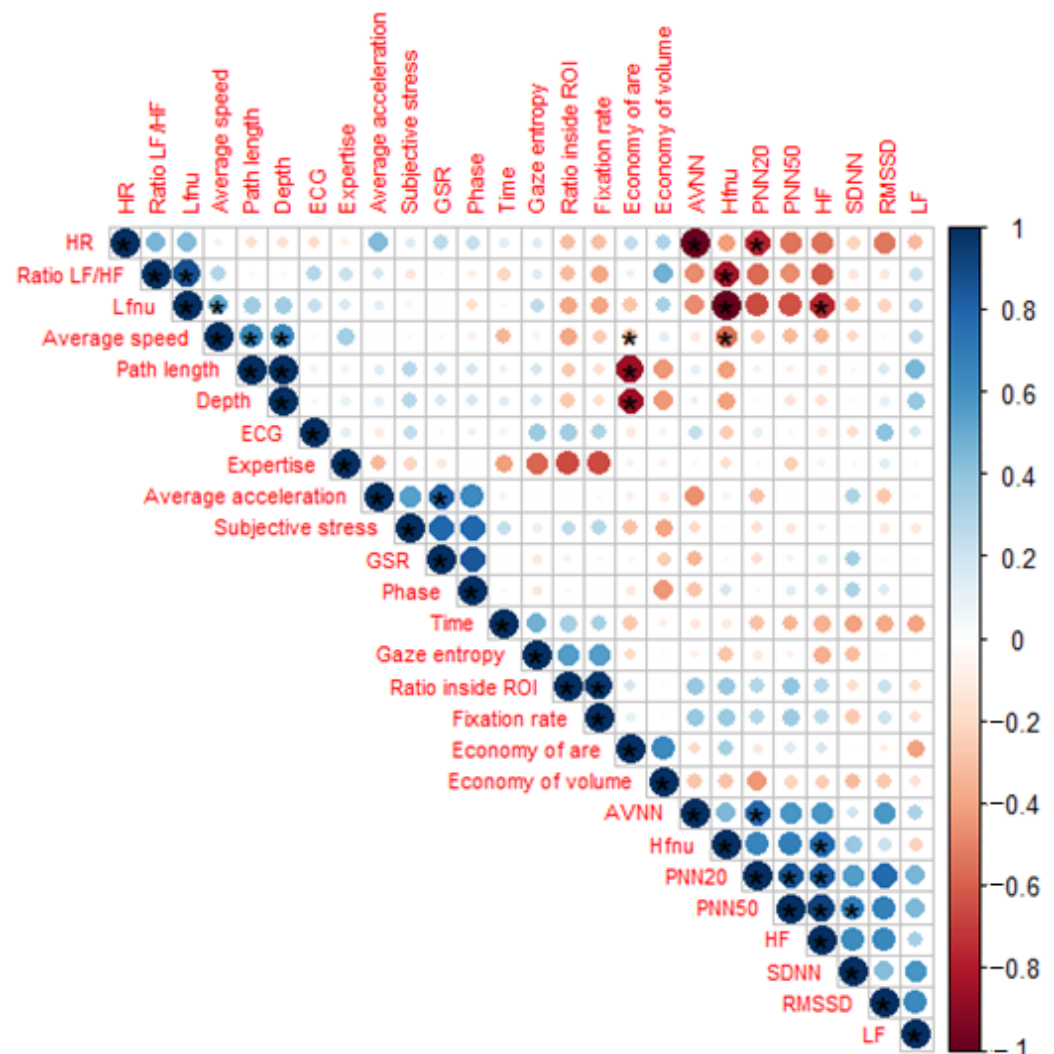


Figure 6. Correlation matrix. Color map and size represent the correlation between the variables. A star is depicted whenever the correlation was statistically significant according to Pearson or Spearman's test.

4. Discussion

In this study, we analyzed 22 different metrics that were obtained during laparoscopic practice: 12 of which were derived from the ECG (in relation with stress), 1 of which was derived from the GSR (in relation with stress), 3 of which were derived from gaze analysis (in relation with visual attention) and 7 of which were derived from the motion tracking analysis of the dominant laparoscopic instrument (in relation with psychomotor skills). Specifically, the analysis was aimed at determining whether (1) stress could be characterized through these metrics, and (2) there was a relation between the metrics of psychomotor skills and personal resourcefulness.

The normalized GSR average value was found to be the most relevant parameter in terms of its relationship with stress, with an increase of 261% between the baseline

and stressful phases of the experiment and a high, significantly positive correlation with the subjective stress rating that was stated by the participants. The relevance of the GSR average values has already been demonstrated in the literature and appears as one of the trending parameters for measuring stress in surgical environments [59–67].

Although the literature suggests that the metrics that are derived from ECG and HRV analyses are highly correlated with stress, none of the time- or frequency-derived metrics seem to have captured the variations in the HR as expected. However, we could not find evidence on how these metrics change with stress (i.e., not all of the subjects showed an increase or decrease in these metrics). This may be due to the fact that a baseline level was not established before the performance of the experiment. This may have led to unwanted bias in the results; for instance, the consumption of caffeine may have led to differences in HR [95].

The average acceleration of the surgical instrument increased significantly during the stressful phase of the experiment (by 15.09%). However, the overall time did not decrease significantly, which suggests that the movements of the instruments in the stressful phase were more abrupt than in the previous phase of the experiment.

The fact that we were able to correlate the GSR (which is a stress parameter) to acceleration indicates that there is a relation between psychomotor skills and personal resourcefulness skills. Other studies have correlated basic psychomotor skills metrics (e.g., the number of errors, time of completion, etc.) with stress parameters through subjective stress scales (e.g., STAI, NASA-TLX). We were able to do so in a quantitative manner and using technical metrics (i.e., acceleration). However, since only the GSR was found to be correlated to stress and, in light of the low power that was yielded in the case of ECG-derived metrics, further studies are necessary to confirm or refute this hypothesis.

The reason why we selected the initial sample size is partly based on the numbers used in similar studies (e.g., Tomizawa et al. [30] used the data from 4 participants and Jones et al. [50] used the data from 6 participants), which are generally low due to the difficulties in the recruitment of surgeons (and even more so in the case of expert surgeons), essentially based on their tight schedules. However, the aftermath of COVID-19 reduced surgeons' availability even more and the safety limitations that were imposed made it difficult to recruit more surgeons. Thus, we ended up with a lower number of surgeons than initially calculated, which turned out to be the greatest limitation of this study.

It is due to this low number of surgeons that we could not compute an in-depth statistical analysis in order to find differences between the experts and nonexperts. Nevertheless, we did carry out a descriptive analysis, finding that, during the stressful phase, two of the three metrics that were derived from the gaze analysis were found to be higher in the nonexpert surgeons as compared to the experts during the stressful phase of the experiment (i.e., the ratio of fixations inside the ROI and the fixation rate). This is consistent with the literature, where fixation analysis has demonstrated to be in close relation to surgical expertise (and even suggested as a possible means towards training) [22,96,97]. This may be indicative of a lower cognitive load in the experts; due to the internalization of knowledge and surgical practice, experts are able to focus not only on the main surgical area but on other areas of the field of view. However, recognizing the low number of experts that were included in this study as one of its greatest limitations, we propose to further explore this in future research with higher participation and complete statistical analysis.

Additionally, none of the psychomotor metrics were found to be influenced by expertise. This may be due to the simplicity of the task (i.e., the task was so simple that there were little to no variations between experts and nonexperts) [98]. Thus, implementing a more difficult laparoscopic task in order to increase the cognitive load (such as suturing) while including stressors could lead to greater variations and better characterization of the differences between expertise levels using the same metrics as in this experiment. This could also be due to the lack of participants to compare experts and nonexperts, for which we highlight again the importance of conducting further experiments with more participants,

once proven feasible that the measurements are valid and that some of the parameters that were monitored effectively measure surgical stress (i.e., the GSR and acceleration).

5. Conclusions

This study stems from the need to find objective assessment methodologies for personal resourcefulness skills in surgical education, while researching the relation between psychomotor and personal resourcefulness skills. To address this need, we proposed to analyze the influences between a set of metrics that are related to stress, gaze and instrument position and how they are affected by stressors and expertise.

After analyzing several metrics, we can conclude (1) that both the mean of the normalized GSR signal and the average acceleration of the surgical instrument are closely related to stress (i.e., they change significantly when presented with stressors) and (2) that the average acceleration was found to be correlated to the GSR, suggesting that there may be a link between psychomotor skills and personal resourcefulness. Further experiments are nevertheless required in order to extract further evidence from other metrics and tasks, together with an in-depth analysis of the influence of expertise.

Thus, the next steps, based on the results of this study, would be aimed at gathering information from surgeons with different specialties while performing a more complex laparoscopic task (such as suturing) in which stressors are included. For these new experiments, we aim at obtaining a larger number of participants in order to ensure higher statistical power. In addition to this, the differences between the obtained metrics in immersive simulated environments would also be of interest. We will also study the validity of the use of these metrics as a part of the assessment of personal resourcefulness as compared with validated scales (such as NOTECHS [99]). Furthermore, the study of a combination of metrics that are closely related to expertise could also contribute to the assessment of skills (both psychomotor and personal resourcefulness, simultaneously).

In conclusion, the results that we obtained demonstrate that it is feasible to use physiological monitoring and gaze analysis in order to obtain relevant information about the stress and attention levels of laparoscopic surgeons, which are directly related to their personal resourcefulness skills. This information could ultimately be used for the assessment of skills that are related to personal resourcefulness in surgical educational settings. Moreover, certain metrics that are related to personal resourcefulness have demonstrated to correlate with psychomotor skills metrics, which opens the way towards a complete objective assessment of surgical proficiency.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/s22030837/s1>, Table S1. Summary of results of extracted metrics as related to stress. All technical metrics presented are from the right instrument (dominant hand). $F(1,14)$ represents the F value of the test, being 1 and 14 the degrees of freedom of the test; $U(8,8)$ represents the U test statistic, being 8 the sample size of subjects in the first and second phase; p indicates the p -value of the corresponding statistical test.

Author Contributions: Conceptualization, C.G.-G., P.S.-G., J.A.S.M., N.S., J.C.R., F.M.S.M., E.J.G. and I.O.; Data curation, C.G.-G., J.A.S.M., J.C.R., F.M.S.M. and I.O.; Formal analysis, C.G.-G., P.S.-G., J.A.S.M., N.S., J.C.R., F.M.S.M., E.J.G. and I.O.; Funding acquisition, P.S.-G. and E.J.G.; Investigation, C.G.-G., P.S.-G., J.A.S.M., N.S., J.C.R., F.M.S.M., E.J.G. and I.O.; Methodology, C.G.-G., J.A.S.M., N.S., J.C.R., E.J.G. and I.O.; Project administration, P.S.-G., F.M.S.M. and E.J.G.; Resources, C.G.-G., P.S.-G., J.A.S.M. and J.C.R.; Software, C.G.-G., N.S. and I.O.; Supervision, P.S.-G., J.A.S.M., E.J.G. and I.O.; Validation, C.G.-G., E.J.G. and I.O.; Visualization, C.G.-G. and N.S.; Writing—original draft, C.G.-G. and N.S.; Writing—review & editing, C.G.-G., P.S.-G., J.A.S.M., J.C.R., F.M.S.M., E.J.G. and I.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the Biomedical engineering and Telemedicine Center that is associated with the Universidad Politécnica de Madrid, Madrid, Spain and the EC-funded Erasmus+ program, grant number [621668-EPP-1-2020-1-ES-EPPKA2-KA].

Institutional Review Board Statement: Ethical review and approval were waived for this study, due to its minimal risk and educational nature.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data available on request due to restrictions eg privacy or ethical.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Rutherford, D.N.; D'Angelo, A.L.D.; Law, K.E.; Pugh, C.M. Advanced Engineering Technology for Measuring Performance. *Surg. Clin. N. Am.* **2015**, *95*, 813–826. [[CrossRef](#)] [[PubMed](#)]
- Yule, S.; Flin, R.; Paterson-Brown, S.; Maran, N. Non-technical skills for surgeons in the operating room: A review of the literature. *Surgery* **2006**, *139*, 140–149. [[CrossRef](#)]
- Agha, R.A.; Fowler, A.J.; Sevdalis, N. The role of non-technical skills in surgery. *Ann. Med. Surg.* **2015**, *4*, 422–427. [[CrossRef](#)]
- O'Neil, H.F.; Perez, R.S.; Baker, E.L. *Teaching and Measuring Cognitive Readiness*; Springer: Boston, MA, USA, 2014; ISBN 9781461475798.
- Pugh, C.M.; DaRosa, D.A. Use of Cognitive Task Analysis to Guide the Development of Performance-Based Assessments for IntraOperative Decision Making. *Mil. Med.* **2013**, *170*, 22–27. [[CrossRef](#)] [[PubMed](#)]
- Wiegmann, D.A.; ElBardissi, A.W.; Dearani, J.A.; Daly, R.C.; Sundt, T.M. Disruptions in surgical flow and their relationship to surgical errors: An exploratory investigation. *Surgery* **2007**, *142*, 658–665. [[CrossRef](#)]
- Dedy, N.J.; Fecso, A.B.; Szasz, P.; Bonrath, E.M.; Grantcharov, T.P. Implementation of an effective strategy for teaching nontechnical skills in the operating room. A single-blinded nonrandomized trial. *Ann. Surg.* **2016**, *263*, 937–941. [[CrossRef](#)]
- Anton, N.E.; Bean, E.A.; Hammonds, S.C.; Stefanidis, D. Application of mental skills training in surgery: A review of its effectiveness and proposed next steps. *J. Laparoendosc. Adv. Surg. Tech.* **2017**, *27*, 459–469. [[CrossRef](#)]
- Datta, V.; Chang, A.; Mackay, S.; Darzi, A. The relationship between motion analysis and surgical technical assessments. *Am. J. Surg.* **2002**, *184*, 70–73. [[CrossRef](#)]
- Dosis, A.; Aggarwal, R.; Bello, F.; Moorthy, K.; Munz, Y.; Gillies, D.; Darzi, A. Synchronized video and motion analysis for the assessment of procedures in the operating theater. *Arch. Surg.* **2005**, *140*, 293–299. [[CrossRef](#)]
- Chmarra, M.K.; Bakker, N.H.; Grimbergen, C.A.; Dankelman, J. TrEndo, a device for tracking minimally invasive surgical instruments in training setups. *Sens. Actuators A Phys.* **2006**, *126*, 328–334. [[CrossRef](#)]
- Oropesa, I.; Sánchez-González, P.; Chmarra, M.K.; Lamata, P.; Fernández, Á.; Sánchez-Margallo, J.A.; Jansen, F.W.; Dankelman, J.; Sánchez-Margallo, F.M.; Gómez, E.J. EVA: Laparoscopic instrument tracking based on endoscopic video analysis for psychomotor skills assessment. *Surg. Endosc.* **2013**, *27*, 1029–1039. [[CrossRef](#)] [[PubMed](#)]
- Chmarra, M.K.; Klein, S.; De Winter, J.C.F.; Jansen, F.W.; Dankelman, J. Objective classification of residents based on their psychomotor laparoscopic skills. *Surg. Endosc.* **2010**, *24*, 1031–1039. [[CrossRef](#)]
- Fried, G.M.; Feldman, L.S. Objective assessment of technical performance. *World J. Surg.* **2008**, *32*, 156–160. [[CrossRef](#)] [[PubMed](#)]
- Li, M.M.; George, J. A systematic review of low-cost laparoscopic simulators. *Surg. Endosc.* **2017**, *31*, 38–48. [[CrossRef](#)] [[PubMed](#)]
- Bell, R.H. Why Johnny cannot operate. *Surgery* **2009**, *146*, 533–542. [[CrossRef](#)] [[PubMed](#)]
- Sharma, B.; Mishra, A.; Aggarwal, R.; Grantcharov, T.P. Non-technical skills assessment in surgery. *Surg. Oncol.* **2011**, *20*, 169–177. [[CrossRef](#)]
- Norman, G.R.; Grierson, L.E.M.; Sherbino, J.; Hamstra, S.J.; Schmidt, H.G.; Mamede, S. Expertise in Medicine and Surgery. In *The Cambridge Handbook of Expertise and Expert Performance*; Cambridge University Press: Cambridge, UK, 2018; pp. 331–355.
- Brunckhorst, O.; Shahid, S.; Aydin, A.; McIlhenny, C.; Khan, S.; Raza, S.J.; Sahai, A.; Brewin, J.; Bello, F.; Kneebone, R.; et al. Simulation-based ureteroscopy skills training curriculum with integration of technical and non-technical skills: A randomised controlled trial. *Surg. Endosc.* **2015**, *29*, 2728–2735. [[CrossRef](#)] [[PubMed](#)]
- Madani, A.; Vassiliou, M.C.; Watanabe, Y.; Al-Halabi, B.; Al-Rowais, M.S.; Deckelbaum, D.L.; Fried, G.M.; Feldman, L.S. What Are the Principles That Guide Behaviors in the Operating Room? *Ann. Surg.* **2017**, *265*, 255–267. [[CrossRef](#)]
- Stefanidis, D.; Anton, N.E.; Howley, L.D.; Bean, E.; Yurco, A.; Pimentel, M.E.; Davis, C.K. Effectiveness of a comprehensive mental skills curriculum in enhancing surgical performance: Results of a randomized controlled trial. *Am. J. Surg.* **2017**, *213*, 318–324. [[CrossRef](#)]
- Law, B.; Lomax, A.J.; Atkins, M.S.; Mackenzie, C.L.; Kirkpatrick, A.E. Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. In *Proceedings of the 2004 Symposium on Eye Tracking Research and Applications*, San Antonio, TX, USA, 22–24 March 2004; pp. 41–48.
- Wilson, M.; McGrath, J.; Vine, S.; Brewer, J.; Defriend, D.; Masters, R. Psychomotor control in a virtual laparoscopic surgery training environment: Gaze control parameters differentiate novices from experts. *Surg. Endosc.* **2010**, *24*, 2458–2464. [[CrossRef](#)]
- Wilson, M.R.; Vine, S.J.; Bright, E.; Masters, R.S.W.; Defriend, D.; McGrath, J.S. Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: A randomized, controlled study. *Surg. Endosc.* **2011**, *25*, 3737–3739. [[CrossRef](#)]

25. Jacob, R.J.K.; Karn, K.S. Eye Tracking in Human-Computer Interaction and Usability Research. Ready to Deliver the Promises. In *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*; Elsevier: Amsterdam, The Netherlands, 2003; pp. 573–605. ISBN 9780080518923.
26. Zhang, Q.; Wang, Z.; Chi, J.; Zhang, P.; Yang, Y. Design and calibration for gaze tracking system. In Proceedings of the IEEE 2nd International Conference on Information Management and Engineering: ICIME'10, Chengdu, China, 16–18 April 2010; pp. 16–18.
27. Land, R.; Meyer, J.H.F. The Scalpel and the 'Mask': Threshold Concepts and Surgical Education. In *Surgical Education*; Fry, H., Kneebone, R., Eds.; Springer: Dordrecht, The Netherlands, 2011; pp. 91–106.
28. Richstone, L.; Schwartz, M.J.; Seideman, C.; Cadeddu, J.; Marshall, S.; Kavoussi, L.R. Eye metrics as an objective assessment of surgical skill. *Ann. Surg.* **2010**, *252*, 177–182. [[CrossRef](#)]
29. Koh, R.Y.I.; Park, T.; Wickens, C.D.; Ong, L.T.; Chia, S.N. Differences in Attentional Strategies by Novice and Experienced Operating Theatre Scrub Nurses. *J. Exp. Psychol. Appl.* **2011**, *17*, 233–246. [[CrossRef](#)]
30. Tomizawa, Y.; Aoki, H.; Suzuki, S.; Matayoshi, T.; Yozu, R. Eye-tracking analysis of skilled performance in clinical extracorporeal circulation. *J. Artif. Organs* **2012**, *15*, 146–157. [[CrossRef](#)]
31. Erridge, S.; Ashraf, H.; Purkayastha, S.; Darzi, A.; Sodergren, M.H. Comparison of gaze behaviour of trainee and experienced surgeons during laparoscopic gastric bypass. *Br. J. Surg.* **2018**, *105*, 287–294. [[CrossRef](#)] [[PubMed](#)]
32. Roach, V.A.; Fraser, G.M.; Kryklywy, J.H.; Mitchell, D.G.V.; Wilson, T.D. Time limits in testing: An analysis of eye movements and visual attention in spatial problem solving. *Anat. Sci. Educ.* **2017**, *10*, 528–537. [[CrossRef](#)] [[PubMed](#)]
33. Sundelin, T.; Lekander, M.; Kecklund, G.; Van Someren, E.J.W.; Olsson, A.; Axelsson, J. Cues of Fatigue: Effects of Sleep Deprivation on Facial Appearance. *Sleep* **2013**, *36*, 1355–1360. [[CrossRef](#)] [[PubMed](#)]
34. Gupta, S.; Mittal, S. Yawning and its physiological significance. *Int. J. Appl. Basic Med. Res.* **2013**, *3*, 11–15. [[CrossRef](#)]
35. Schleicher, R.; Galley, N.; Briest, S.; Galley, L. Blinks and saccades as indicators of fatigue in sleepiness warnings: Looking tired? *Ergonomics* **2008**, *51*, 982–1010. [[CrossRef](#)]
36. Arora, S.; Sevdalis, N.; Nestel, D.; Woloshynowych, M.; Darzi, A.; Kneebone, R. The impact of stress on surgical performance: A systematic review of the literature. *Surgery* **2010**, *147*, 318–330. [[CrossRef](#)]
37. Wetzel, C.M.; Kneebone, R.L.; Woloshynowych, M.; Nestel, D.; Moorthy, K.; Kidd, J.; Darzi, A. The effects of stress on surgical performance. *Am. J. Surg.* **2006**, *191*, 5–10. [[CrossRef](#)]
38. Crewther, B.T.; Shetty, K.; Jarchi, D.; Selvadurai, S.; Cook, C.J.; Leff, D.R.; Darzi, A.; Yang, G.Z. Skill acquisition and stress adaptations following laparoscopic surgery training and detraining in novice surgeons. *Surg. Endosc.* **2016**, *30*, 2961–2968. [[CrossRef](#)] [[PubMed](#)]
39. LaPorta, A.J.; McKee, J.; Hoang, T.; Horst, A.; McBeth, P.; Gillman, L.M.; Kirkpatrick, A.W. Stress Inoculation: Preparing Outside the Box in Surgical Resuscitation and Education. *Curr. Trauma Rep.* **2017**, *3*, 135–143. [[CrossRef](#)]
40. Abe, T.; Dar, F.; Amnatrakul, P.; Aydin, A.; Raison, N.; Shinohara, N.; Khan, M.S.; Ahmed, K.; Dasgupta, P. The effect of repeated full immersion simulation training in ureterorenoscopy on mental workload of novice operators. *BMC Med. Educ.* **2019**, *19*, 39. [[CrossRef](#)]
41. Causer, J.; Vickers, J.N.; Snelgrove, R.; Arsenault, G.; Harvey, A. Performing under pressure: Quiet eye training improves surgical knot-tying performance. *Surgery* **2014**, *156*, 1089–1096. [[CrossRef](#)] [[PubMed](#)]
42. Allen, R.; Robinson, A.; Allen, S.; Nathan, E.; Coghlan, E.; Leung, Y. Designing meditation for doctor well-being: Can 'Om' help obstetrics and gynaecology doctors? *Australas. Psychiatry* **2020**, *28*, 342–347. [[CrossRef](#)]
43. Anton, N.E.; Bean, E.A.; Myers, E.; Stefanidis, D. Optimizing learner engagement during mental skills training: A pilot study of small group vs. individualized training. *Am. J. Surg.* **2020**, *219*, 335–339. [[CrossRef](#)]
44. Stefanidis, D.; Anton, N.E.; McRary, G.; Howley, L.D.; Pimentel, M.; Davis, C.; Yurco, A.M.; Sevdalis, N.; Brown, C. Implementation results of a novel comprehensive mental skills curriculum during simulator training. *Am. J. Surg.* **2017**, *213*, 353–361. [[CrossRef](#)]
45. Timberlake, M.D.; Stefanidis, D.; Gardner, A.K. Examining the impact of surgical coaching on trainee physiologic response and basic skill acquisition. *Surg. Endosc.* **2018**, *302*, 4183–4190. [[CrossRef](#)]
46. Goldberg, M.B.; Mazzei, M.; Maher, Z.; Fish, J.H.; Milner, R.; Yu, D.; Goldberg, A.J. Optimizing performance through stress training—An educational strategy for surgical residents. *Am. J. Surg.* **2018**, *216*, 618–623. [[CrossRef](#)]
47. Anton, N.E.; Beane, J.; Yurco, A.M.; Howley, L.D.; Bean, E.; Myers, E.M.; Stefanidis, D. Mental skills training effectively minimizes operative performance deterioration under stressful conditions: Results of a randomized controlled study. *Am. J. Surg.* **2018**, *215*, 214–221. [[CrossRef](#)]
48. Lemaire, J.B.; Wallace, J.E.; Lewin, A.M.; de Grood, J.; Schaefer, J.P. The effect of a biofeedback-based stress management tool on physician stress: A randomized controlled clinical trial. *Open Med.* **2011**, *5*, e156–e163.
49. Alobid, I.; De Pablo, J.; Mullol, J.; Centellas, S.; Parramon, G.; Carrasco, J.; Armario, A.; Bernal-Sprekelsen, M. Increased cardiovascular and anxiety outcomes but not endocrine biomarkers of stress during performance of endoscopic sinus surgery: A pilot study among novice surgeons. *Arch. Otolaryngol.-Head Neck Surg.* **2011**, *137*, 487–492. [[CrossRef](#)] [[PubMed](#)]
50. Jones, K.I.; Amawi, F.; Bhalla, A.; Peacock, O.; Williams, J.P.; Lund, J.N. Assessing surgeon stress when operating using heart rate variability and the State Trait Anxiety Inventory: Will surgery be the death of us? *Colorectal Dis.* **2015**, *17*, 335–341. [[CrossRef](#)] [[PubMed](#)]
51. Böhm, B.; Rötting, N.; Schwenk, W.; Grebe, S.; Mansmann, U. A prospective randomized trial on heart rate variability of the surgical team during laparoscopic and conventional sigmoid resection. *Arch. Surg.* **2001**, *136*, 305–310. [[CrossRef](#)]

52. Marrelli, M.; Gentile, S.; Palmieri, F.; Paduano, F.; Tatullo, M. Correlation between surgeon's experience, surgery complexity and the alteration of stress related physiological parameters. *PLoS ONE* **2014**, *9*, e112444. [[CrossRef](#)]
53. Weenk, M.; Alken, A.P.B.; Engelen, L.J.L.P.G.; Bredie, S.J.H.; van de Belt, T.H.; van Goor, H. Stress measurement in surgeons and residents using a smart patch. *Am. J. Surg.* **2018**, *216*, 361–368. [[CrossRef](#)]
54. Grantcharov, P.D.; Boillat, T.; Elkabany, S.; Wac, K.; Rivas, H. Acute mental stress and surgical performance. *BJS Open* **2019**, *3*, 119–125. [[CrossRef](#)]
55. Dishman, R.K.; Nakamura, Y.; Garcia, M.E.; Thompson, R.W.; Dunn, A.L.; Blair, S.N. Heart rate variability, trait anxiety, and perceived stress among physically fit men and women. *Int. J. Psychophysiol.* **2000**, *37*, 121–133. [[CrossRef](#)]
56. Zhai, J.; Barreto, A. Stress recognition using non-invasive technology. In Proceedings of the FLAIRS 2006 Nineteenth International Florida Artificial Intelligence Research Society Conference, Melbourne Beach, FL, USA, 11–13 May 2006; Volume 1, pp. 395–401.
57. Acharya, U.R.; Joseph, K.P.; Kannathal, N.; Lim, C.M.; Suri, J.S. Heart rate variability: A review. *Med. Biol. Eng. Comput.* **2006**, *44*, 1031–1051. [[CrossRef](#)]
58. Borrego, A.; Latorre, J.; Alcaniz, M.; Llorens, R. Reliability of the Empatica E4 wristband to measure electrodermal activity to emotional stimuli. In Proceedings of the International Conference on Virtual Rehabilitation, ICVR, Tel Aviv, Israel, 21–24 July 2019.
59. Smith, W.D.; Chung, Y.H.; Berguer, R. A virtual instrument ergonomics workstation for measuring the mental workload of performing video-endoscopic surgery. *Stud. Health Technol. Inform.* **2000**, *70*, 309–315. [[CrossRef](#)]
60. Hoedemaeker, M. *Summary Description of Workload Indicators: WP1 Workload Measures*; University of Leeds: Leeds, UK, 2002.
61. Ershad, M.; Rege, R.; Fey, A.M. Meaningful Assessment of Robotic Surgical Style using the Wisdom of Crowds. *Int. J. Comput. Assist. Radiol. Surg.* **2018**, *13*, 1037–1048. [[CrossRef](#)] [[PubMed](#)]
62. Hurley, A.M.; Kennedy, P.J.; O'Connor, L.; Dinan, T.G.; Cryan, J.F.; Boylan, G.; O'Reilly, B.A. SOS save our surgeons: Stress levels reduced by robotic surgery. *Gynecol. Surg.* **2015**, *12*, 197–206. [[CrossRef](#)]
63. Shi, Y.; Ruiz, N.; Taib, R.; Choi, E.; Chen, F. Galvanic skin response (GSR) as an index of cognitive load. In Proceedings of the Conference on Human Factors in Computing Systems, San Jose, CA, USA, 28 April–3 May 2007; pp. 2651–2656.
64. Schuetz, M.; Gockel, I.; Beardi, J.; Hakman, P.; Dunschede, F.; Moenk, S.; Heinrichs, W.; Junginger, T. Three different types of surgeon-specific stress reactions identified by laparoscopic simulation in a virtual scenario. *Surg. Endosc.* **2008**, *22*, 1263–1267. [[CrossRef](#)]
65. Wilson, C.; Chahine, S.; Cristancho, S.; Aquil, S.; Mandurah, M.; Levine, M.; Sener, A. Unusual suspects: Real-time physiological evaluation of stressors during laparoscopic donor nephrectomy. *Can. Urol. Assoc. J.* **2020**, *15*, 205–209. [[CrossRef](#)]
66. Bakker, J.; Pechenizkiy, M.; Sidorova, N. What's your current stress level? Detection of stress patterns from GSR sensor data. In Proceedings of the IEEE International Conference on Data Mining, ICDM, Vancouver, BC, Canada, 11 December 2011.
67. Georgiou, K.; Larentzakis, A.; Papavassiliou, A.G. Surgeons' and surgical trainees' acute stress in real operations or simulation: A systematic review. *Surgeon* **2017**, *15*, 355–365. [[CrossRef](#)] [[PubMed](#)]
68. Bartolomeo, L.; Lin, Z.; Zecca, M.; Sessa, S.; Ishii, H.; Xu, H.; Uemura, M.; Tomikawa, M.; Hashizume, M.; Takanishi, A. Surface EMG and heartbeat analysis preliminary results in surgical training: Dry boxes and live tissue. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, Boston, MA, USA, 30 August–3 September 2011; pp. 1113–1116.
69. Duru, D.G.; Duru, A.D.; Barkana, D.E.; Sanli, O.; Ozkan, M. Assessment of surgeon's stress level and alertness using EEG during laparoscopic simple nephrectomy. In Proceedings of the International IEEE/EMBS Conference on Neural Engineering, NER, San Diego, CA, USA, 6–8 November 2013; pp. 452–455.
70. Morales, J.M.; Ruiz-Rabelo, J.F.; Diaz-Piedra, C.; Di Stasi, L.L. Detecting Mental Workload in Surgical Teams Using a Wearable Single-Channel Electroencephalographic Device. *J. Surg. Educ.* **2019**, *76*, 1107–1115. [[CrossRef](#)]
71. Al-Shargie, F.; Kiguchi, M.; Badruddin, N.; Dass, S.C.; Hani, A.F.M.; Tang, T.B. Mental stress assessment using simultaneous measurement of EEG and fNIRS. *Biomed. Opt. Express* **2016**. [[CrossRef](#)]
72. Al-Shargie, F.; Tang, T.B.; Kiguchi, M. Assessment of mental stress effects on prefrontal cortical activities using canonical correlation analysis: An fNIRS-EEG study. *Biomed. Opt. Express* **2017**. [[CrossRef](#)]
73. Shetty, K.; Leff, D.R.; Orihuela-Espina, F.; Yang, G.Z.; Darzi, A. Persistent prefrontal engagement despite improvements in laparoscopic technical skill. *JAMA Surg.* **2016**, *151*, 682–684. [[CrossRef](#)]
74. Arora, S.; Sevdalis, N.; Aggarwal, R.; Sirimanna, P.; Darzi, A.; Kneebone, R. Stress impairs psychomotor performance in novice laparoscopic surgeons. *Surg. Endosc.* **2010**, *24*, 2588–2599. [[CrossRef](#)]
75. Arora, S.; Aggarwal, R.; Moran, A.; Sirimanna, P.; Crochet, P.; Darzi, A.; Kneebone, R.; Sevdalis, N. Mental practice: Effective stress management training for novice surgeons. *J. Am. Coll. Surg.* **2011**, *212*, 225–233. [[CrossRef](#)]
76. Maher, Z.; Milner, R.; Cripe, J.; Gaughan, J.; Fish, J.; Goldberg, A.J. Stress training for the surgical resident. *Am. J. Surg.* **2013**, *205*, 169–274. [[CrossRef](#)]
77. Wetzel, C.M.; George, A.; Hanna, G.B.; Athanasiou, T.; Black, S.A.; Kneebone, R.L.; Nestel, D.; Woloshynowych, M. Stress management training for surgeons—a randomized, controlled, intervention study. *Ann. Surg.* **2011**, *253*, 488–494. [[CrossRef](#)] [[PubMed](#)]
78. Krohne, H.W.; De Bruin, J.T.; El-Giamal, M.; Schmukle, S.C. The assessment of surgery-related coping: The coping with surgical stress scale (COSS). *Psychol. Health* **2000**, *15*, 135–159. [[CrossRef](#)]

79. Arora, S.; Tierney, T.; Sevdalis, N.; Aggarwal, R.; Nestel, D.; Woloshynowych, M.; Darzi, A.; Kneebone, R. The imperial stress assessment tool (ISAT): A feasible, reliable and valid approach to measuring stress in the operating room. *World J. Surg.* **2010**, *34*, 1756–1763. [[CrossRef](#)]
80. Wilson, M.S.; Middlebrook, A.; Sutton, C.; Stone, R.; McCloy, R.F. MIST VR: A virtual reality trainer for laparoscopic surgery assesses performance. *Ann. R. Coll. Surg. Engl.* **1997**, *79*, 403–404. [[PubMed](#)]
81. Platte, K.; Alleblas, C.C.J.; Inthout, J.; Nieboer, T.E. Measuring fatigue and stress in laparoscopic surgery: Validity and reliability of the star-track test. *Minim. Invasive Ther. Allied Technol.* **2019**, *28*, 57–64. [[CrossRef](#)]
82. Kildebro, N.; Amirian, I.; Gögenur, I.; Rosenberg, J. Test re-test reliability and construct validity of the star-track test of manual dexterity. *PeerJ* **2015**, *3*, e917. [[CrossRef](#)]
83. Frankenhaeuser, M.; Lundberg, U. The influence of cognitive set on performance and arousal under different noise loads. *Motiv. Emot.* **1977**, *1*, 139–149. [[CrossRef](#)]
84. Wijsman, J.; Grundlehner, B.; Penders, J.; Hermens, H. Trapezius muscle EMG as predictor of mental stress. In Proceedings of the Wireless Health 2010, WH'10, San Diego, CA, USA, 5–6 October 2010; pp. 155–163.
85. Cao, C.G.L.; Zhou, M.; Jones, D.B.; Schwaitzberg, S.D. Can surgeons think and operate with haptics at the same time? *J. Gastrointest. Surg.* **2007**, *11*, 1564–1569. [[CrossRef](#)]
86. Sharma, N.; Gedeon, T. Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Comput. Methods Programs Biomed.* **2012**, *108*, 1287–1301. [[CrossRef](#)] [[PubMed](#)]
87. Analog Devices AD8232 Datasheet. Available online: <https://www.analog.com/media/en/technical-documentation/datasheets/AD8232.pdf> (accessed on 25 March 2020).
88. BangGood CJMCU-6701 Datasheet. Available online: <http://img.banggood.com/file/products/20181017011131CJMCU-6701.pdf> (accessed on 25 March 2020).
89. Pope, B. How to perform 3- or 5-lead monitoring. *Nursing* **2002**. [[CrossRef](#)]
90. Oropesa, I.; Escamiroso, F.P.; Sánchez-Margallo, J.A.; Enciso, S.; Rodríguez-Vila, B.; Martínez, A.M.; Sánchez-Margallo, F.M.; Gómez, E.J.; Sánchez-González, P. Interpretation of motion analysis of laparoscopic instruments based on principal component analysis in box trainer settings. *Surg. Endosc.* **2018**, *32*, 3096–3107. [[CrossRef](#)]
91. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate—A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B-Methodol.* **1995**, *57*, 289–300. [[CrossRef](#)]
92. Erdfelder, E.; Faul, F.; Buchner, A.; Lang, A.G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods* **2009**, *41*, 1149–1160. [[CrossRef](#)]
93. Faul, F.; Erdfelder, E.; Lang, A.G.; Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **2007**, *39*, 175–191. [[CrossRef](#)] [[PubMed](#)]
94. Cohen, J. *Statistical Power Analysis for the Behavioural Science*; Academic Press: Cambridge, MA, USA, 1988; ISBN 0805802835.
95. Lane, J.D.; Pieper, C.F.; Phillips-Bute, B.G.; Bryant, J.E.; Kuhn, C.M. Caffeine affects cardiovascular and neuroendocrine activation at work and home. *Psychosom. Med.* **2002**, *64*, 595–603. [[CrossRef](#)]
96. Chen, H.E.; Sonntag, C.C.; Pepley, D.F.; Prabhu, R.S.; Han, D.C.; Moore, J.Z.; Miller, S.R. Looks can be deceiving: Gaze pattern differences between novices and experts during placement of central lines. *Am. J. Surg.* **2019**, *217*, 362–367. [[CrossRef](#)]
97. Marín-Conesa, E.; Sánchez-Ferrer, F.; Grima-Murcia, M.D.; Sánchez-Ferrer, M.L. The Application of a System of Eye Tracking in Laparoscopic Surgery: A New Didactic Tool to Visual Instructions. *Front. Surg.* **2021**, *8*, 643611. [[CrossRef](#)]
98. Oropesa, I.; Sánchez-González, P.; Lamata, P.; Chmarra, M.K.; Pagador, J.B.; Sánchez-Margallo, J.A.; Sánchez-Margallo, F.M.; Gómez, E.J. Methods and tools for objective assessment of psychomotor skills in laparoscopic surgery. *J. Surg. Res.* **2011**, *171*, 81–95. [[CrossRef](#)] [[PubMed](#)]
99. Flin, R.; Martin, L.; Goeters, K.M.; Hormann, H.J.; Amalberti, R.; Valot, C.; Nijhuis, H. Development of the NOTECHS (non-technical skills) system for assessing pilots' CRM skills. *Hum. Factors Aerosp. Saf.* **2003**, *3*, 97–119. [[CrossRef](#)]