Citation for published version:

de Arriba Pérez, F., Santos-Gago, J. M., Caeiro-Rodríguez, M., Fernández Iglesias, M. J. Evaluation of Commercial-Off-The-Shelf Wrist Wearables to Estimate Stress on Students. *J. Vis. Exp.* (136), e57590, doi: 10.3791/57590 (2018)

Accepted Manuscript

Link to published version: https://dx.doi.org/10.3791/57590

TITLE:

Evaluation of Commercial-Off-The-Shelf Wrist Wearables to Estimate Stress on Students

AUTHORS & AFFILIATIONS:

Francisco de Arriba-Pérez[1], Juan M. Santos-Gago[1], Manuel Caeiro-Rodríguez[1], Manuel J. Fernández Iglesias[1]

[1] Department of Telematics Engineering, University of Vigo, Campus Lagoas-Marcosende, Vigo, Galicia, Spain

Corresponding Author:
Francisco de Arriba-Pérez
Email Address: farriba@uvigo.es
Tel: +34 986 814 073

Email Addresses of Co-authors:

Juan M. Santos-Gago	(Juan.Santos@det.uvigo.es)
Manuel Caeiro-Rodríguez	(Manuel.Caeiro@det.uvigo.es)
Manuel J. Fernández Iglesias	(manolo@det.uvigo.es)

KEYWORDS:

wrist-wearables, quantification, multimodal analytics, stress detection, machine learning, e-learning.

SHORT ABSTRACT:

A protocol to evaluate solutions based on commercial-off-the-shelf (COTS) wrist wearables to estimate stress in students is proposed. The protocol is carried out in two phases, an initial laboratory-based stress induction test, and a monitoring stage taking place in the classroom while the student is performing academic activities.

LONG ABSTRACT:

Wearable commercial-off-the-shelf (COTS) devices have become popular during the last years to monitor sports activities, primarily among young people. These devices include sensors to gather data on physiological signals such as heart rate, skin temperature or galvanic skin response. By applying data analytics techniques to these kinds of signals, it is possible to obtain estimations of higher-level aspects of human behavior. In the literature, there are several works describing the use of physiological data collected using clinical devices to obtain information on sleep patterns or stress. However, it is still an open question whether data captured using COTS wrist wearables is sufficient to characterize the learners' psychological state in educational settings. This paper discusses a protocol to evaluate stress estimation from data obtained using COTS wrist wearables. The protocol is carried out in two phases. The first stage consists of a controlled laboratory experiment, where a mobile app is used to induce different stress levels in a student by means of a relaxing video, a Stroop Color and Word test, a Paced Auditory Serial Addition test, and a hyperventilation test. The second phase is carried out in the classroom,

where stress is analyzed while performing several academic activities, namely attending to theoretical lectures, doing exercises and other individual activities, and taking short tests and exams. In both cases, both quantitative data obtained from COTS wrist wearables and qualitative data gathered by means of questionnaires are considered. This protocol involves a simple and consistent method with a stress induction app and questionnaires, requiring a limited participation of support staff.

INTRODUCTION:

State-of-the-art wearable technologies are widely available, and their application environments are continuously expanding. We can find in the market many different devices, among which COTS wrist wearables[1], such as smart watches and smart bands, are popular among athletes as a personal physical fitness monitoring tool[2]. By applying data analytic techniques, the data obtained using these devices can be processed to provide indicators such as general physical state, sleep quality or recovery factor. The demonstrated applicability in this area raised interest in the academic community about their possible application to other fields, especially in the health domain[3,4], although the strict requirements of clinical trials limit their introduction. However, in a less demanding context such as education, we can find in the literature recent investigations involving the use of different types of wearable devices, both related to teaching activities[5,6] and to the estimation of certain characteristics of the student such as sleep patterns[7], or the analysis of students' engagement in different educational activities[8].

In our case, we focus on analyzing COTS wrist wearable devices as means to collect physiological signals that would eventually facilitate stress estimation, which in turn is a key aspect in educational contexts. Stress has a relevant influence in the development of academic activities and overall students' performance. For example, stress levels are directly related to the onset of the burnout syndrome in students[9–11], and high stress levels are especially relevant during the freshman year, where drop-out rates between 20% and 30%[12,13] are common. Detecting and controlling stress indicators could dramatically improve academic performance.

The use of COTS wrist wearable devices is justified because they have sensors that provide information on physiological signals that have been widely used by the scientific community in stress assessment and detection. Some of the signals referred to in the literature used for this purpose include heart rate (HR)[14], heart rate varaibility[15], skin temperature (ST)[16], respiration[14], and galvanic skin response (GSR)[17]. These signals can be collected by COTS wrist wearables. However, they do not offer the same performance as clinical devices. There are differences related to the accuracy of sensors among devices[18–21]. Nevertheless, previous works[18–21] have shown that, in a slow movement scenario, COTS wrist wearable sensors have error patterns similar to specialized devices.

The aim of this paper is to introduce a protocol to evaluate different solutions for stress estimation in students using COTS wrist wearables. There are many arrangements that can be proposed to estimate stress levels, involving the use of different wrist wearable devices and data analytics techniques, and more specifically machine learning algorithms. COTS wrist wearables are characterized by their high fragmentation, heterogeneity and interoperability problems[22]. Three companies have an aggregated market share of almost 50%[23], but many other companies account for much smaller individual market shares, with an aggregated share above 50%. On the other hand, in terms of heterogeneity, not all wearables have the same number and type of sensors, with accelerometers and h sensors being the most common, and ST's and GSR's being only present in 5% of the devices studied. As for interoperability, there are different operating systems and data collection approaches that are not compatible with each other. As for the machine learning techniques that can be applied to estimate stress from the data collected by means of a wrist device, there are many options available[24], including decision trees, neural networks, nearest neighbor approaches, Naïve Bayes classifiers, *etc.* To sum up, there is a great variety of solutions that may be developed for stress estimation, so it is instrumental to design an evaluation protocol to facilitate the comparison among different tentative options to eventually select the most suitable in a given context.

For the implementation of the protocol, several tools are needed (**Figure 1**). First, a COTS wrist wearable device is needed to fetch physiological data. This wearable device should have at least HR monitoring capabilities, but additional sensors are desirable (*e.g.,* accelerometer, ST, GSR sensors). Second, a smartphone running the PhysiologicalSignal app is required to collect the data captured by the wearable device. Third, a tablet running the StressTest app is needed to run stress induction exercises (the smartphone could be used instead the tablet for this purpose). Fourth, some questionnaires to collect qualitative data on students' perception on stress. Fifth, a server with a Web service[25] to perform data collection and pre-processing, and a Web dashboard to show the evolution of the signals. And finally, a data analytics package[26] to process the data collected about students using machine learning techniques.

The evaluation protocol is organized into two phases. The first one, the laboratory phase, is carried out in a comfortable room, where different stress levels (*i.e.*, "relax", "concentrated stress" and "stress") are induced to a target subject (a student) through several common stress-inducing tasks. The second part takes place in the classroom, and it involves monitoring the student during the accomplishment of several academic activities: theoretical explanations, individual activities, short tests, exams, *etc.* During the implementation of this protocol, the subject's physiological signals are captured by means of a wrist device. Finally, these signals are processed by machine learning algorithms to provide estimations on the level of stress.

During the laboratory phase, the StressTest app is used to induce different stress levels. This app guides the subject to the completion of four different tasks. The first task is to create a baseline for stress analysis. In this task, the student visualizes a 4-minute relaxing video in which different shots of a sunset on a bridge are shown. The second task is an adaptation of the Stroop Color and Word Test[27] (SCWT). Every two seconds, the subject must choose the color in which the name of a color is painted (red, green, orange, blue and purple). Several buttons located at the bottom of the screen containing the initial letter of each color are available for the subject to choose the painted color at each time. For example, the button that refers to blue depicts the letter B. In our case, this test is divided into three different levels of difficulty. For the first level (SCWT1), the colored "words of colors" will appear in the same order as the buttons, so color and name match directly. This level is taken as baseline, as it does not involve any difficulty and the subject should only press the buttons properly, always in the same order. For the second level (SCWT2), the colored "words of colors" appear randomly, but the correspondence between name and color is maintained. Every time the subject fails a beep is emitted, and if two errors are made, the correct color score will be reset. For the last, most difficult level (SCWT3), name and color do not match. In this way this level is intended to be more complex and stressful for the subject. The third task consists on the Paced Auditory Serial Addition test (PASAT)[28], which measures how the student experiences a concentration test. During this task, a sequence of consecutive numbers is played aloud, and the student must add the last two numbers and write the result in the provided on-screen box before listening to the next number. In this task, if the subject makes a mistake, a disturbing event occurs to generate stress (two numbers sound at the same time or a long period of silence in maintained). In this case, if three errors are committed, the sum account will be reset. The fourth task consists on a hyperventilation activity to induce the same variation in the physiological signals that would provoke a stressful situation[17]. At the end of each task and level, the subject has to indicate the level of perceived stress, using the application itself, according to a 5-value Likert scale.

During the classroom phase, students carry out their ordinary academic activities together with the rest of their classmates. The protocol focuses on the stress levels that occur during classroom-specific activities. At the end of the lecture, a brief questionnaire (Annex 1) is completed by the student to indicate the perceived level of stress in the several activities according to a 5-value scale.

PROTOCOL:

All methods described below have been approved by the regional government of Galicia's committee for research ethics of Pontevedra-Vigo-Ourense (reg. code 2017/336). The protocol was implemented for first year students at the School of Telecommunication Engineering - University of Vigo, both in a comfortable laboratory room and in several lectures and practice sessions of a bachelor's degree course on Computer Architectures.

1. Prepare the Devices

1.1. Connect the smartphone and tablet device to a stable internet connection.

1.2. Turn on Bluetooth communications in the smartphone.

1.3. In the smartphone, search in the corresponding official app store the wrist wearable application. Download and install it.

1.4. In the smartphone, search for the PhysiologicalSignals app to capture physiological signals. Download and install it.

Note: Currently, the app is a beta version and access can be provided by request.

1.5. In the tablet, search for the StressTest app to be used in the research laboratory experiments. Download and install it.

Note: Currently, the app is a beta version and access can be provided by request.

1.6. Turn on the COTS wrist wearable device and place the wearable.

1.7. In the smartphone, open the official COTS wrist wearable application.

Note: The app will proceed to synchronize the wearable device with the smartphone. In some devices, an e-mail address is required.

1.8. In the smartphone, open the PhysiologicalSignals app.

1.8.1. In case of being notified of a sensor access request, accept it.

1.8.2. Check the device. Wait for the PhysiologicalSignals app to display the word **Weared** in green.

Note: This indicate that the wearable device has been detected and therefore the transmission of information from the sensors to the smartphone is enabled. If this message does not appear repeat from step 1.6.

2. The Laboratory Phase

2.1. Prepare the laboratory setting. Choose a comfortable and non-disturbing room without distracting noise and with a comfortable temperature (between 22 °C and 26 °C).

2.2. Turn on the wrist wearable device, place it around the subject´s non-dominant wrist and place the headphones on the head of the student. Fit the wearable tightly but comfortably around the wrist.

2.3. Connect the smartphone and tablet to a stable internet connection and verify that the Bluetooth connection is active.

2.4. In the smartphone, launch the PhysiologicalSignals app.

2.4.1. Wait for the app to display the word **Weared** in green.

2.4.2. Select the **Change User** option in the left configuration menu and provide the ID of the subject who will complete the tests and click **Save**.

2.5. In a laptop, access the dashboard and enter the test administrator’s ID and password.

Note: Currently, for private and security concerns, access to the dashboard is only available under request.

2.5.1. Select the subject ID and the subject's stress tab.

2.5.2. Check the physiological signals evolution and wait for the wearable device to reach thermal stability before starting the experiment.

Note: The thermal stability is identified as a plateau in the graph.

2.6. In the tablet, launch the StressTest application.

2.6.1. Explain to the subject the four laboratory tasks. Show some of the screens and actions to perform during each one of the tasks.

Note: This is very important, because the subject should feel stressed or relaxed in accordance to the performed activities, and not fear or concern about what is going to happen.

2.6.2. Tell the student not to rest their arms on the table and to use the hand where the wearable device is placed to perform the activities.

2.6.3. Enter the same user ID as in step 2.4.2 and click the arrow.

2.7. Launch the video task and give full control to the student.

2.7.1. Observe that the task is carried out without incident.

2.7.2. When the task is finished, check that the subject provides the perceived stress.

2.8. Launch the Stroop Color task (SCWT) consecutively for levels 1, 2 and 3.

2.8.1. For each level, observe that the subtask is carried out without incident.

2.8.2. When each subtask is finished, check that the subject provides the perceived stress.

2.8.3. Only for level 3 and only in case the subject does not solve it after 4 minutes, terminate the task by pressing the arrow located at the top of the screen.

2.9. Launch the Paced Auditory Serial Addition test (PASAT).

2.9.1. Observe that the task is carried out without incident.

2.9.2. In case the subject does not solve the PASAT test after 4 minutes, terminate the task by pressing the arrow located at the top of the screen.

2.9.3. When the task is finished, check that the subject provides the perceived stress.

2.10. Launch the Hyperventilation test.

2.10.1. Observe the evolution of HR using the dashboard. If physiological signals do not change significantly, ask the subject to increase inspiration and expiration rates gradually.

2.10.2. In case the subject feels dizziness or uncomfortable halt this task. In any case, complete the task after four minutes.

2.10.3. When the task is finished, check that the subject provides the perceived stress.

3. The Classroom Phase

3.1. Turn on the wrist wearable device and place the wearable around the subject´s non-dominant wrist. Fit the wearable tightly but comfortably around the wrist.

3.2. Connect the smartphone to a stable internet connection and verify the Bluetooth connection is active.

3.3. In the smartphone, launch the PhysiologicalSignals app.

3.3.1. Wait for the app to display the word **Weared** in green.

3.3.2. Select in the configuration menu the **Change User** option, provide the ID of the subject who will complete the tests and click **Save**.

3.4. In a laptop, access the dashboard and enter the test administrator's ID and password.

3.4.1. Select the subject ID and the subject's stress tab.

3.4.2. Check the evolution of physiological signals.

3.5. Take annotations about any relevant event occurring in the classroom in relation to the student-teacher interaction.

Note: Relevant information and basic events will be used to label physiological samples afterwards. Example events are a question from the teacher to the student, or a theoretical explanation is initiated.

3.6. At the end of the lecture, ask the subject to complete the questionnaire about their level of stress at specific times during the session, according to a 5-level scale.

4. Data Analysis

4.1. In a laptop, access the dashboard and enter the test administrator's ID and password.

4.1.1. Select the subject ID and the subject's stress tab.

4.1.2. Select the day of a classroom experiment.

4.2. Label the samples of the subject by identifying activities and perceived stress levels.

4.2.1. Identify lecture-room activities and their duration according to the starting and finishing times and their types.

4.2.2. For each activity, select a perceived stress level.

4.3. For each subject and each session, download the file with the tagged samples.

Note: A comma-separated-values (CSV) file is created for each student, each row reflecting the values of the physiological signals with their standard deviation, slope and diff, the activity type, the activity-based stress (*i.e.*, the stress associated by default to the activity) and the subject perceived stress.

4.4. Launch the data analytics package.

4.4.1. Choose a set of classifiers (*e.g.*, SVM, C4.5, k-NN, Random Forest, Naïve Bayes and Zero R) and import the CSV file for all students for each session.

4.4.2. Train and evaluate classifiers using the 10-fold cross-validation technique.

Note: Depending on the analyses, activity type, activity-based stress or stress perceived, shall be selected as dependent variable for the analysis.

4.4.3. Finally, check the results for accuracy and error rates.

REPRESENTATIVE RESULTS:

The protocol discussed was put into practice in a Computer Architectures course in the first year of the Telecommunication Engineering degree at the University of Vigo. This course has more than 200 students enrolled who are organized into 10 working groups. To carry out this experiment, students from four of the groups were invited to enroll at the beginning of the academic year. The project attracted considerable interest among the students, and around 30 students volunteered to participate in the study. From them, 12 students were randomly selected for participation.

The COTS wrist wearable device selected for our experiments has HR, ST, GSR and accelerometer sensors. The choice of this wearable was based on its variety of sensors and the provision of real-time data feeding. Technical conditions in which sensor data is collected were also taken into account. Data capture is performed at certain frequencies, generally imposed by the operation of the sensors, but also due to the device's energy-saving characteristics. In the case of the selected device, HR was sampled every second (1 Hz). The accelerometer offered 62 Hz, 31 Hz and 8 Hz as sampling frequencies, from which 8 Hz was selected because it offers enough granularity for movement capture with reasonable energy requirements when compared to the other frequencies. GSR may be sampled at 0.2 or 5 Hz. In this case, we opted to gather GSR data once every 5 seconds. As for the accelerometer, this frequency provided enough granularity while keeping energy requirements to a minimum. Finally, ST is sampled at the same frequency as HR (*i.e.*, 1 Hz). Data collected by the device is transferred to the PhysiologicalSignals app in the smartphone every second, including the HR and ST sample, the maximum acceleration value, and the last value for GSR collected. To reduce HR noise, the server applies to the received data a FIR filter commonly used in real-time applications[29] and in the filtering of ECG signals[30], using a 15-sample window.

Information gathered during laboratory and classroom sessions is stored in the server´s database. This information should be downloaded to be processed using a data analytics package. The set of generated data files contains raw signals' data and variables derived from those signals. More specifically, for each raw physiological signal (HR, ST, GSR and accelerometer), its standard deviation (st), slope (sl), and the difference between the present value and the extreme value in the last 30 seconds are recorded.

The laboratory phase of the protocol was carried out in a comfortable room of the Telematics Engineering department that has the appropriate conditions for the experiment. **Figure 2** depicts the evolution of HR, GSR and ST values collected during one of these sessions for an actual student. As can be seen in the figure, significant variations in the physiological signals occur as the student performs each of the tasks (video, STC1, STC2, STC3, PASAT and Hyperventilation) included in the experiment. A relatively high initial HR value can be observed, most probably due to the stress induced when facing this task for the first time while being monitored. The rapid growth of ST during the hyperventilation test is also noteworthy.

Also observed during the laboratory experiments were the remarkable variations in the physiological signals at specific experimental moments, no matter that these periods were not always perceived as stressful by the target student. This is due to the fact that perceived stress is a subjective variable, and participating students do not fully agree in a common concept of stress. During the laboratory phase, it was intended to generate brief periods of high stress. These brief periods of stress were sometimes defined as frustration, but not as stress, which leads participating students to respond differently to what their physiological signals expressed. This effect can be visualized in the graphs in **Figure 3**. For example, in the interval between

12:15 and 12:20 (completion of the last test of the Stroop Color and Word Test) the strong GSR variations are a clear symptom of potential stress. These strong variations are also present between 12:25 and the end of the test (Hyperventilation test), but on both occasions, the user claimed to feel a similarly low stress level.

The situation discussed above stresses the subjective character of stress evaluation in such a short period of time. As a consequence, from the candidates for dependent variables in data sets (*i.e.*, activity type, activity-based stress, or subject-perceived stress) we opted for activity-based stress. This variable defines stress levels according to the level of difficulty of the task addressed and not on the answers provided by the students about their perceived stress levels at the end of each task. This way, video watching would be tagged as “relax” while SCWT3 and PASAT would be labelled as “concentration” and the Hyperventilation test as “stress”. Note that samples from SCWT1 and SCWT2 were discarded in our case because in a previous pilot research was observed that, on average, SCWT1 and SCWT2 are activities that show a transition between a relaxed feeling (reached during video visualization) and stressful one. For these reason, we discarded from our analysis the signals from these 2 activities, and we included only those from video visualization, SCWT3, PASAT and Hyperventilation activities. The HR, ST and GSR variations among these states (relax, concentration, stress) are summarized in **Figure 4**. This figure depicts the physiological signal quartiles for the three stress levels in the 12 students involved in the experiment. In general, HR and GSR signals gradually increase as the student faces tasks of increasing difficulty. Also, in all cases the temperature level is affected. However, in some cases it increases for relaxed events and decreases in stressful situations, while in other cases it occurs just the opposite depending of the person.

In order to analyze the correlation observed visually in the variation of the physiological signals, machine learning techniques were applied over processed CSV files. To avoid initial transitory variations for each task and level, only the last 3 minutes of each activity are considered in order to avoid non-representative samples. In particular, several classification algorithms, particularly SVM, C4.5, k-NN, Random Forest, NaiveBayes and ZeroR, were trained to detect stress situations from the collected physiological signals. The trained classifiers became high accuracy, low mean absolute error rates and high Cohen’s Kappa index level stress detectors, as it is shown in **Table 1**. For all the 12 subjects and algorithms (except ZeroR), the accuracy of stress detection in over 90%, mean absolute error value is near 0 and Cohen’s kappa index is close to 1.

The classroom phase defined in the protocol took place during actual course sessions in the lecture rooms of the School of Telecommunications Engineering. Several academic activities were considered for this study: theoretical lectures; questions arbitrarily asked by the teacher to the students about some aspect of the course; doubts or questions posed to the teacher by students; short tests; regular exams/finals consisting of collection of problems to be solved by the student in 50-70 minutes.

The visualization of the evolution of physiological signals in this case shows that variations are subtler, that is, the differences in signal values for different activities are smaller than during the laboratory phase. The most relevant variations were observed during classroom sessions in which a regular lecture occurs after a pop quiz is completed. In this case, one or several of the physiological signals suffer significant differences, as illustrated in **Figure 5**. This figure depicts the signals captured for a student facing a short test (first part of the graphs). During the test, the most relevant variable would be HR. It can be observed that the student has a higher heart rate when compared to theoretical lecture time. In the same way, skin temperature is kept relatively low when compared to theoretical lecture time, when it raises around 1 °C.

To analyze this in a numerical way, the correlation between the variations in the signals and the activities addressed by the students, machine learning techniques were applied analogously to the laboratory phase. The results for the combined pop quiz and lecture sessions show an average classification accuracy of 97.62% (± 3.82) using C4.5. Note that for the analysis of these sessions skin temperature was discarded due to possible biases in the final result. During the transition period between the pop quiz and the following lecture students leave the classroom for approximately 20 minutes, with dramatically affects temperature values.

A comprehensive formal analysis of the collected classroom sessions is still in progress. This is a complex process where several challenging situations are addressed. First, abrupt short-time variations in the physiological signals are frequently observed with no associated stress-generating event. In most cases, these periods last for less than one minute without anything significant being recorded by the researcher. Another incidence observed is the instability of the GSR values when the wearable is not well adjusted or if sudden movements occur. Both situations result in a very low GSR values, close to 0 µS. In a similar way, although much less usual, there are incorrect ST values, close to the ambient temperature, when the wearable is too big for the wrist of the user and therefore is loosely worn. To eliminate the analysis errors derived from these situations, affected variables are discarded. Note that all the signals monitored can be candidates to detect stress situations and different classifiers may be trained using different combinations of signals, but anomalous values would compromise classification no matter the classifier selected.

FIGURE AND TABLE LEGENDS:

Figure 1. Tools used in the proposed protocol. This figure represents all the elements involved in the protocol and their interactions.

Figure 2. Stress variation in a laboratory session. This figure shows the different parts in which the laboratory protocol is divided. Each part presents a clear variation in the physiological signals.

Figure 3. Stress variation perceived for a student in a laboratory session. This figure shows the discrepancies between the strong variations of the physiological signals of a student during a laboratory session and their answer to the stress quiz.

Figure 4. Physiological signal percentiles for 12 students participating in a laboratory session. This figure represents a percentile summary for each subject. The strong physiological signal variations between each stress situation can be visualized.

Figure 5. HR, ST and GSR variations during classroom activities. Physiological signals variation during a short test (Left). Physiological signals variation during a theoretical lecture (Right).

Table 1. Accuracy, mean absolute error, and Cohen's Kappa index values obtained for SVM, C4.5, k-NN, Random Forest, NaiveBayes and ZeroR machine learning classifiers using data from the 12 students participating in the laboratory experiment.

DISCUSSION:

COTS wearable devices are among the most popular consumer electronics products available today. These devices are typically used to monitor physical activities, but their capabilities and performance could be of great interest in other areas. In this paper, a protocol to evaluate the use of COTS wearable devices for estimating stress in learning environments is discussed. The definition of such a protocol is especially relevant in order to analyze different solutions involving wearables and machine learning algorithms. The protocol is intended to be used in educational settings, where the validation of stress detection procedures and their eventual introduction may provide significate benefits. For example the use of wearable devices can contribute to reduce the high levels of stress associated to the so-called burnout syndrome[9–11], and as a consequence the dropout rate at universities[12,13], while improving academic performance.

A critical aspect to consider is the Bluetooth link between the wearable and the smartphone. This wireless connection between both devices may be broken during the test, so it is necessary to pay special attention to it through the visualization of the data collected in the dashboard. Although recovery is performed automatically after a short period of time (*i.e.*, an interval ranging from 1 to 10 minutes), this interruption may cause the loss of the samples in that interval. To reduce the amount of information lost, it may be convenient to manually reset the smartphone device. Other aspect to be considered is the initial skin temperature sensor value, as it may affect the achievement of skin stability, which may be delayed up to 10 minutes.

The main advantages of the protocol proposed in this research are its applicability to a large group of students, its minimal need for support using automated mobile apps, its simplicity in the preparation of the devices involved in the experiment and its low intrusiveness while carrying out the classroom phase. This protocol provides a fast and simple method applicable in controlled environments, such as classrooms or university laboratories. Besides, technological abilities of participating students are not an issue, as the protocol is based in straightforward technical concepts understandable by an average university student independently of their academic field. As stated in the literature[31], reproducibility in experimental sciences requires a thorough and clear description of the protocols applied and the results thereof. The protocol discussed in this paper has been designed in a modular way according to simple, straightforward steps, which facilitates the reproduction of the experiments discussed and their extension[32]. Among the most relevant design aspects facilitating reproducibility, we can name the conciseness of the laboratory phase and its automated implementation by means of standalone mobile apps. Additionally, the classroom phase does not require any interaction with the students beyond academic activities. Most students pointed out the simplicity of the process, and no complaints were reported in relation to their involvement in the experiments. To sum up, collected evidence so far indicates that this protocol may be applied to subjects with a broader profile and in fields different to education, such as health facilities or the working place. Besides, this protocol offers the possibility to study several machine learning solutions with which to test the best algorithms to implement depending on the requirements of the experiments and on the wearable device selected. The use of applications to induce stress and to provide a dashboard to display and tag samples facilitates the training of custom stress models in a single laboratory session.

The main limitations of the proposed solution are related on the subjects' variability and the reproducibility of academic activities. Recreating exactly the same conditions and situations taking place in lecture sessions is practically impossible. On the other hand, the stress experienced by each student is very personal, as in general there are different responses to the same stimuli. In addition, there are hardware-related issues related to the wearable devices themselves, such as different access methods, different sensors, access to physiological signals in real time, or battery life. These technical requirements restrict eligible wearables to a limited range of devices. In our case, eligible devices include those compatible with smart Bluetooth capabilities and smart bands with a SDK compatible with major SO smartphone devices. The number of compatible devices is expected to increase along the next years.

The proposed protocol is intended to serve as an instrument to eventually define richer student models than those presently used in learning management systems or student information systems. For example, the new information captured with the wearable device according to the protocol discussed could be applied to the early detection of situations affecting performance such as fatigue or stress, and to guide students to overcome these situations. An alternative to this protocol may be based on wearable devices worn also outside the classroom in order to detect variations in physiological signals over a longer period of time. This approach involves several challenges, such as a constantly changing ambient temperature, or the subject under study being forced to always be close to their smartphone to prevent data loss. Finally, this protocol may be also applied to other courses and educational levels, which would facilitate the capture of additional evidence on how stress influences academic performance for students with different skills or fields of study.

ACKNOWLEDGMENTS:
This work is supported by the Spanish State Research Agency and the European Regional Development Fund (ERDF) under the PALLAS (TIN2016-80515-R AEI/EFRD, EU) project.

DISCLOSURES:
The authors have nothing to disclose.

REFERENCES:

1. IDC IDC Forecasts Wearables Shipments to Reach 213.6 Million Units Worldwide in 2020 with Watches and Wristbands Driving Volume While Clothing and Eyewear Gain Traction. at <http://www.idc.com/getdoc.jsp?containerId=prUS41530816> (2016).

2. Piwek, L., Ellis, D. A., Andrews, S. & Joinson, A. The rise of consumer health wearables:promises and barriers. *PLoS Medicine* **13** (2) (2016).

3. Rudner, J. *et al.* Interrogation of Patient Smartphone Activity Tracker to Assist Arrhythmia Management. *Annals of Emergency Medicine* **68** (3), 292–294, doi:10.1016/j.annemergmed.2016.02.039 (2016).

4. Gao, Y., Li, H. & Luo, Y. An empirical study of wearable technology acceptance in healthcare. *Industrial Management & Data Systems* **115** (9), 1704–1723, doi:10.1108/IMDS-03-2015-0087 (2015).

5. Lukowicz, P. *et al.* Glass-physics: using google glass to support high school physics experiments. *Proceedings of the 2015 ACM International Symposium on Wearable Computers - ISWC '15*, 151–154, doi:10.1145/2802083.2808407 (2015).

6. Sapargaliyev, D. Wearables in Education: Expectations and Disappointments. 73–78, doi:10.1007/978-3-662-48978-9_7 (2015).

7. de Arriba Pérez, F., Caeiro Rodríguez, M. & Santos Gago, J. M. How do you sleep? Using off the shelf wrist wearables to estimate sleep quality, sleepiness level, chronotype and sleep regularity indicators. *Journal of Ambient Intelligence and Humanized Computing*, 1–21, doi:10.1007/s12652-017-0477-5 (2017).

8. Espinosa, H. G., Lee, J., Keogh, J., Grigg, J. & James, D. A. On the Use of Inertial Sensors in Educational Engagement Activities. *Procedia Engineering* **112**, 262–266, doi:10.1016/j.proeng.2015.07.242 (2015).

9. Travers, C. J. & Cooper, C. L. *El Estrés de los profesores : la presión en la actividad docente*. at <https://dialnet.unirioja.es/servlet/libro?codigo=110437> (Paidós: 1997).

10. Maslach, C. & Jackson, S. E. The measurement of experienced burnout*. *Journal of occupational Behavior* **2**, 99–113 (1981).

11. Maslach, C., Jackson, S. & Leiter, M. Maslach Burnout Inventory. Palo Alto. (1986).

12. Kitsantas, A., Winsler, A. & Huie, F. Self-Regulation and Ability Predictors of Academic Success During College: A Predictive Validity Study. *Journal of Advanced Academics* **20**, doi:10.4219/jaa-2008-867 (2008).

13. Deberard, C., Scott, M., Glen, I., Spielmans, D. C. & Julka Predictors of academic achievement and retention among college freshmen: a longitudinal study. *College Student Journal* **381**, 66–80 (2004).

14. Healey & Jennifer Anne Wearable and automotive systems for affect recognition from physiology. (2000).

15. Vrijkotte, T. G. M., van Doornen, L. J. P. & de Geus, E. J. C. Effects of Work Stress on Ambulatory Blood Pressure, Heart Rate, and Heart Rate Variability. *Hypertension* **35** (4) (2000).

16. Karthikeyan, P., Murugappan, M. & Yaacob, S. Descriptive Analysis of Skin Temperature Variability of Sympathetic Nervous System Activity in Stress. *Journal of Physical Therapy Science* **24** (12), 1341–1344, doi:10.1589/jpts.24.1341 (2012).

17. Santos Sierra, A. de Design, implementation and evaluation of an unconstrained and contactless biometric system based on hand geometry and stress detection. (2012).

18. Natale, V., Drejak, M. & Erbacci, A. Monitoring sleep with a smartphone accelerometer. *Sleep and Biological Rhythms* at <http://onlinelibrary.wiley.com/doi/10.1111/j.1479-8425.2012.00575.x/full> (2012).

19. Guo, F., Li, Y., Kankanhalli, M. & Brown, M. An evaluation of wearable activity monitoring devices. *Proceedings of the 1st ACM international workshop on Personal data meets distributed multimedia* at <http://dl.acm.org/citation.cfm?id=2512882> (2013).

20. Wallen, M. P. *et al.* Accuracy of Heart Rate Watches: Implications for Weight Management. *PLOS ONE* **11** (5), e0154420, doi:10.1371/journal.pone.0154420 (2016).

21. Wang, R. *et al.* Accuracy of Wrist-Worn Heart Rate Monitors. *JAMA Cardiology* **2** (1), 104, doi:10.1001/jamacardio.2016.3340 (2017).

22. de Arriba Pérez, F., Caeiro Rodríguez, M. & Santos Gago, J. M. Collection and Processing of Data from Wrist Wearable Devices in Heterogeneous and Multiple-User Scenarios. *Sensors* **16** (9), 1538, doi:10.3390/s16091538 (2016).

23. IDC Wearables Aren't Dead, They're Just Shifting Focus as the Market Grows 16.9% in the Fourth Quarter, According to IDC. at <https://www.idc.com/getdoc.jsp?containerId=prUS42342317> (2017).

24. Mark, H., Ian, W. & Eibe, F. *Data Mining: Practical Machine Learning Tools and Techniques*. (Morgan Kaufmann Publishers: 2011).

25. de Arriba Pérez, F., Santos Gago, J. M. & Caeiro Rodríguez, M. Analytics of biometric data from wearable devices to support teaching and learning activities. *Journal of Information Systems Engineering & Management* **1**, 41–54, doi:10.20897/lectito.201608 (2016).

26. Machine Learning Group at the University of Waikato Weka 3 - Data Mining with Open Source Machine Learning Software in Java. at <https://www.cs.waikato.ac.nz/mL/weka/>

27. Zhai, J. & Barreto, A. Stress Detection in Computer Users Based on Digital Signal Processing of Noninvasive Physiological Variables. *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, 1355–1358, doi:10.1109/IEMBS.2006.259421 (2006).

28. Tombaugh, T. N. A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). *Archives of Clinical Neuropsychology* **21** (1), 53–76, doi:10.1016/j.acn.2005.07.006 (2006).

29. Qinru Fan & Yigang Wang The real-time realization of filtering of speech with DSP TMS320VC5416 Chip. *2010 International Conference on Educational and Information Technology*, doi:10.1109/ICEIT.2010.5608388 (2010).

30. González Barajas, J. E., Velandia Cárdenas, C. & Nieto Camacho, J. Implementación de filtro digital en tiempo real para detección de la onda R. *Revista Tecno Lógicas* **18** (34), 75–86 (2015).

31. Mesirov, J. P. Computer science. Accessible reproducible research. *Science (New York, N.Y.)* **327** (5964), 415–6, doi:10.1126/science.1179653 (2010).

32. American Journal Experts How to Write an Easily Reproducible Protocol. *2018* at <https://www.aje.com/en/arc/how-to-write-an-easily-reproducible-protocol/>