A Critique of Electrodermal Activity Practices at CHI

Ebrahim Babaei
ebabaei@student.unimelb.edu.au
The University of Melbourne
Melbourne, Australia

Benjamin Tag
benjamin.tag@unimelb.edu.au
The University of Melbourne
Melbourne, Australia

Tilman Dingler
tilman.dingler@unimelb.edu.au
The University of Melbourne
Melbourne, Australia

Eduardo Velloso
evelloso@unimelb.edu.au
The University of Melbourne
Melbourne, Australia

ABSTRACT

Electrodermal activity data is widely used in HCI to capture rich and unbiased signals. Results from related fields, however, have suggested several methodological issues that can arise when practices do not follow established standards. In this paper, we present a systematic methodological review of CHI papers involving the use of EDA data according to best practices from the field of psychophysiology, where standards are well-established and mature. We found severe issues in our sample at all stages of the research process. To ensure the validity of future research, we highlight pitfalls and offer directions for how to improve community standards.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**.

KEYWORDS

Electrodermal Activity, Galvanic Skin Response, EDA, GSR

ACM Reference Format:
Ebrahim Babaei, Benjamin Tag, Tilman Dingler, and Eduardo Velloso. 2021. A Critique of Electrodermal Activity Practices at CHI. *ACM Trans. Graph.* 37, 4, Article 111 (August 2021), 14 pages. https://doi.org/10.1145/3411764.3445370

1 INTRODUCTION

As an indicator of sympathetic nervous system activity and one of the most popular response systems in psychophysiology [80], electrodermal activity (EDA) has become a widely used way of operationalising various psychological constructs including arousal, attention, and stress. EDA—also known as Galvanic Skin Response (GSR)—refers to *"changes in electrical activity of palmar and plantar skin"* [13] and is measured through electrodes on the surface of the skin of specific body parts. The idea behind measuring EDA is that the sympathetic nervous system activates the sweat glands in these body parts, giving researchers a glimpse into otherwise invisible states [22].

The potential offered by EDA has not gone unnoticed by the HCI community—its low cost, relatively unobtrusive form factor, and sensitivity to psychological processes make it a very attractive element in the user experience toolbelt [26]. HCI researchers have employed EDA in numerous studies to detect changes in psychological and mental states relevant to the study of the interaction between people and technology, with contributions including novel sensing devices, user experience studies, and artistic installations.

However, despite the promises of rich insights, pitfalls abound in handling EDA data. A tempting view of EDA is that it is solely controlled by the sympathetic nervous system and that it cannot be intentionally altered or overwritten, serving as an objective ground truth of emotional experience. This makes EDA a probe for continuously measuring individuals' experiences sidestepping the limitations of typical self-report methods such as bias, deception, and interruption. However, EDA data has critical limitations that are often overlooked. For example, in the context of emotion research, EDA data can offer insights about the intensity of an emotion (arousal), but not about whether it is positive or negative (valence). Despite its ostensible simplicity, recording and analyzing EDA signals is a delicate and often intricate undertaking, involving knowledge from psychology, physiology, and signal processing. The underlying psycho-physiological mechanism of EDA is complex and not well-understood, leaving various factors unnoticed that potentially affect the signal. The conductance of the skin is affected by several psychological processes, creating the risk of confounds in its interpretation. This risk is further amplified by HCI's drive and ability to conduct studies in everyday settings, outside the controlled confines of the lab [12]. Finally, it is often difficult to pinpoint exactly which was the stimulus that led to a skin conductance response, due to the latency between the appearance of the stimulus and the associated response.

This context creates a severe risk for HCI researchers to build their research upon a field outside their expertise, threatening the validity of their results. We take this as an opportunity to improve community practices involving EDA data, echoing recent calls for more rigour in HCI methods [20, 21, 96].

In this paper, we analyse current practices in ACM CHI Conference on Human Factors in Computing Systems (CHI) papers

Authors' addresses: Ebrahim Babaei, ebabaei@student.unimelb.edu.au, The University of Melbourne, Melbourne, Australia; Benjamin Tag, benjamin.tag@unimelb.edu.au, The University of Melbourne, Melbourne, Australia; Tilman Dingler, tilman.dingler@unimelb.edu.au, The University of Melbourne, Melbourne, Australia; Eduardo Velloso, evelloso@unimelb.edu.au, The University of Melbourne, Melbourne, Australia.

0730-0301/2021/8-ART111 $15.00
https://doi.org/10.1145/3411764.3445370

involving EDA data, highlighting severe issues, and offering advice for future work. We conducted a systematic literature review, analysing 41 CHI papers as to how well they implement standard best practices from the field of psychophysiology, where the analysis of EDA data is well-established and mature. We found severe issues in our sample, with problems in the reporting, design, collection, and analysis. To assist the community in preventing these mistakes in the future and to generally improve the quality of CHI research involving EDA data, we highlight pitfalls and offer directions for how to improve community standards. This paper contributes to better science and the wider and more systematic use of EDA to measure psychological states and processes in a continuous, direct, and unbiased manner.

2 BACKGROUND

In this section, we explain the biological underpinnings of electrodermal activity and the technical aspects of the analysis of an EDA signal. This works as a primer for readers with little familiarity with the topic and clarifies the terminology we use throughout the paper.

2.1 Physiological foundations

Electrodermal activity (EDA) refers to "the variation of the electrical properties of the skin" [9], which is a direct consequence of dermal sweat gland activity [23]. These changes are due to the high quantity of electrolytes in sweat, leading to an increased conductance of the skin with increased perspiration [77].

The main purpose of perspiration is the regulation of the core body temperature through evaporation, which is carried out by eccrine sweat glands [11]. Even though most eccrine sweat glands are involved in regulating the body's temperature (thermoregulation) [26], plantar (on the foot soles) and palmar (on the hand palms) sweat glands are activated by the sympathetic nervous system [25]. Increases in sweat gland activity in these areas are typically elicited by psychological and emotional states [12]—also called "emotional sweating". Further, these areas show the highest concentration of eccrine glands on the human body [49, 77], which results in a pronounced response to emotional stimuli [26, 29]. As such, increased sympathetic arousal, therefore, prompts an increased number of ducts to fill up with sweat. This results in a lower electrical resistance in that set of filled ducts, producing a measurable change in the EDA signal [77]. In other words, *the activity of the sympathetic nervous system, which controls the body's fight or flight responses, is directly related to changes in the level of the EDA signal.*

2.2 EDA signal

Two categories of methods have been proven effective in measuring the EDA signal: *exosomatic* and *endosomatic* [91]. The more widely used *exosomatic approach* uses an externally applied constant voltage source that is connected to the body via electrodes. The external source can either use a direct current (DC) or an alternating current (AC) [13]. Using DC, the device measures the skin resistance or conductance; using AC, the device measures the skin admittance or impedance. The *endosomatic approach* measures the skin potential directly without applying any voltage [91]. It requires an electrode on an active site (e.g. palm) and one reference

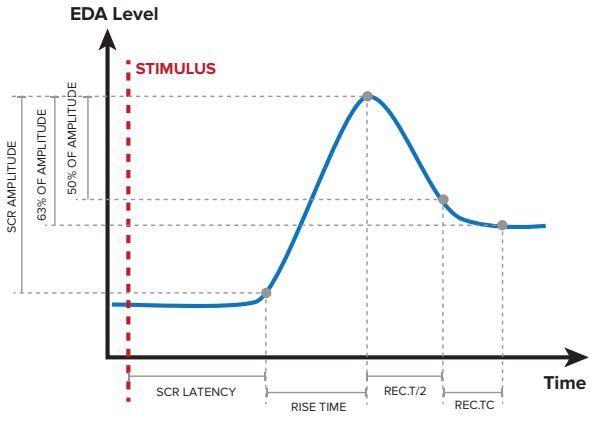

Figure 1: Shape of a skin conductance response (SCR) and its parameters

electrode on an inactive site (e.g. forearm) of body. In contrast to the exosomatic method which results in different signals depends on the voltage or current applied, the endosomatic method results in only one signal, the "skin potential".

Both methods produce a signal that can be decomposed in two, namely the *phasic* and *tonic* components (see Figure 2). The *phasic* component reflects spontaneous responses to an external stimulus and is characterised by high-frequency spikes in the signal [30]. These are often triggered by eliciting stimuli, or sometimes by sources not controlled by the experimenters [91]. This component is commonly analysed through its response amplitude (magnitude, degree of reaction), and the number of responses triggered (frequency, habituation time) [30, 49, 91].

The *tonic* component reflects the emotional arousal baseline and is characterised by a rather inertial and slow response to changes in the participant's condition, such as cognitive stress or workload [30]. The tonic EDA component usually changes over a longer period of time, such as over the length of a full experiment. Each tonic and phasic components of EDA are usually operationalised into *parameters*, such as the conductance level and non-specific response frequency (for the tonic component) and rise time, recovery time, and amplitude (for the phasic component). These parameters are summarised in Table 1.

2.3 Terminology

In early studies in psychology, EDA was mostly referred to as GSR (Galvanic Skin Reflex or Response). However, this term is no longer recommended due to its simplistic assumption that the skin is a galvanic element and the misconception of considering any EDA responses to be a reflex [12]. More recently, the term EDA increased in popularity and is the recommended standard [13].

Another issue found in early psychology work was that authors used inconsistent terms for the parameters of the tonic and phasic components of the EDA signal, such as "amplitude", "delay", and "recovery time". Therefore, to establish a consistent terminology for EDA parameters and help readers to more easily understand publications, Brown [14] developed a set of names, acronyms, and

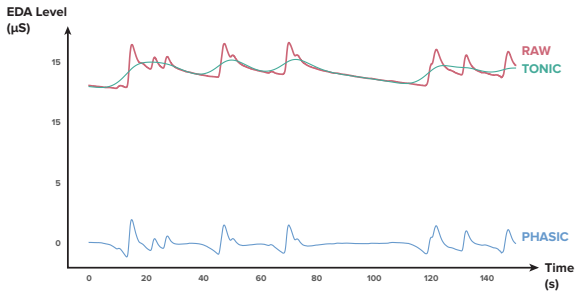

Figure 2: Raw EDA signal, and presentation of *tonic* and *phasic* components of the signal

definitions for EDA signal parameters, which was later amended by Boucsein [12]. According to this terminology, based on what is measured, authors should use the terms **SC** for Skin Conductance, **SR** for Skin Resistance, **SY** for Skin Admittance, **SZ** for Skin Impedance and **SP** for Skin Potential. Boucsein [12] also proposed the general term **ED** (Electrodermal) for when the specific type of measurement is not relevant. In addition, a third letter in this terminology is used to indicate Response (**R**)—referring to changes in the *phasic* component of the EDA signal—or Level (**L**)—which is the value of *tonic* component of EDA. For example, SCR refers to Skin Conductance Response, SPL refers to Skin Potential Level, and EDL refers to Electrodermal Level, which encompasses SCL, SPL, SZL, SYL, or SRL. Most papers in the literature on the topic either use SC (Skin Conductance) directly or transform their signal into SC. Therefore to be consistent within this paper and with the literature, we use SC parameters and definitions throughout this paper.

2.4 EDA parameters

A typical EDA signal exhibits Skin Conductance Responses (SCR) and Skin Conductance Level (SCL), with SCRs typically being more protruding (see Figure 2). When a surprising stimulus is perceived, the *phasic* component of the EDA signal expresses a number of changes in amplitude after the stimulus, resulting in an SCR. Although some authors use the more specific term "Event-Related SCR (ER.SCR)"(e.g. [86, 88]) for reactions to specific stimuli, the term SCR usually refers to responses in the presence of a stimulus. Phasic parameters include the amplitude of the response, the response latency, the rise time until the amplitude is reached, and the recovery time. Figure 1 depicts an SCR and its phasic parameters and Table 1 summarises them.

The *tonic* component of the EDA signal is usually operationalised as the SCL, which can be measured during a non-stimulus rest period and mathematically calculated by taking the mean of a series of measurements. In case an SCR (used for measuring the *phasic* response) cannot be directly linked to a stimulus, it counts as a "non-specific skin conductance response" (NS.SCR) and, as such, is considered a *tonic* measure [77]. The frequency of these events (NS.SCR frequency) is a measure that is often used as a tonic parameter.

2.5 Applications of EDA

Recording EDA data can be relatively inexpensive and unobtrusive compared with physiological signals such as eye-tracking, facial expressions, and thermal imaging, which require a rather cumbersome setup. It is possible to record EDA with little more than a small wearable device, so its convenience and the opportunities it provides to measure different psychological states make it a popular tool in HCI. Below, we review the application domains in which EDA has been explored in CHI papers.

VR: Many CHI papers that have used EDA data were related to Virtual Reality (VR). Because VR headsets cover the eyes and parts of the head and body, it makes it difficult, if not impossible, to use other physiological sensors, such as eye-tracking glasses, facial thermal cameras, and electroencephalography caps—but not EDA. Measuring presence—the sense of "being there"—is a key topic of VR research, and EDA data has the potential to provide insights about it. The idea behind it is that when users feel present, they should exhibit similar reactions as those expected in the corresponding physical environment. For example, Schulz et al. [82] used EDA in a VR climbing environment to measure anxiety and stress and Ogawa et al. [70] measured the sense of body ownership in realistic and non-realistic VR environments. Gromala et al. [43] used EDA in a biofeedback loop for pain management.

VR research can also benefit from EDA data in capturing different aspects of the user experience. Ranasinghe et al. [79] measured users' arousal in different settings of a multi-sensory VR environment and previous works have measured flow [37] and affect [5] with EDA in VR games. Putze et al. [78] measured the impact of using questionnaires inside vs. outside the VR environment on users' presence using EDA.

Games: EDA has been used in games research both to evaluate the player experience and as an input modality for designing game mechanics. The first type of application of EDA in games is to measure elements of the player experience, such as emotion [58] and positive enjoyment [5]. Mirza-Babaei et al. [66] used EDA to improve the game design process. In their approach, a comparison between the EDA signal recorded from players and the EDA desired by the designers informs game designers as to how their games should be modified to match the intended experience.

The second type of application involves the use of EDA by the game itself. Nacke et al. [67] proposed an approach in which players could interact with the game by deliberately trying to affect their EDA signal. Kuikkaniemi et al. [53] used EDA in a biofeedback loop to adapt game mechanics according to players' EDA signal. Even just visualizing the EDA signal within the game environment can be valuable for self-reflection. For example, Wallner et al. [98] proposed a method for visualizing players' EDA signal as an indicator of arousal in games.

Usability and Interface Design: This type of research uses EDA to measure users' experiences with the interface of interest, collecting data about the severity of usability problems and frustration, for example [15]. Use cases range from the evaluation of traditional GUIs, such as websites [39] to more complex cases, such as automotive [35] and olfactory interfaces [39]. EDA has also been used to visualize users' emotional states [63] to detect whether they

Phasic	**SCR**: Skin conductance response. Short-lasting changes in conductivity. **SCR Latency**: Latency between a stimulus and the corresponding change in SCR. **SCR Rise Time**: A rise time from the initial deflection to the peak. **Recovery Time**: Decline in the SC toward the initial value before the SCR. **SCR rec.t/2**: Half-time recovery refers to declining by 50% of the SCR amplitude. **SCR rec.tc**: 63% recovery, which refers to declining by 63% of the SCR amplitude.
Tonic	**SCL**: Skin conductance level in the absence of phasic SCRs. **NS.SCR Frequency**: The frequency of phasic increases in skin conductance that look like stimulus-elicited SCRs but occur in the absence of external stimuli.

Table 1: Definitions of tonic and phasic EDA parameters

perceived notifications [31] and to measure cognitive load [85], comfort, and anxiety while driving [27].

Learning: The analysis of students' EDA signals can be informative about their learning experience. Echeverria et al. [28] and Martinez-Maldonado et al. [60] used EDA to measure the arousal of nursing students during simulations and to communicate it to their supervisors through visualizing this data. Tan et al. [87] also used EDA to measure students' stress and fed this information back to the instructor in video-mediated collaborative learning.

Art: Similarly to its use in games, art projects have used EDA both as part of the artwork and to gain insights about the audience experience. Šimbelis et al. [95] used EDA as an input for a machine to generate a painting based on users' EDA signal. Wang et al. [99] used EDA to measure audience engagement with a live show. Similarly, Latulipe et al. [55] used it to measure engagement with a recorded show and visualized it for artists (e.g. directors, choreographers) to learn more about audience engagement with the show. EDA was also used to improve the communication between an AI and a musician for music improvisation [62].

2.6 Summary

As these examples demonstrate, EDA has wide applicability and rich potential for use by CHI researchers. However, to avoid potential pitfalls and to best leverage this potential, it is necessary to follow best practices for the collection and interpretation of this data. A recent review in the field of consumer behaviour [16] found significant issues in how EDA data was collected, analysed, and reported. Given recent calls for increased transparency in reporting [20] and for caution in interpreting theories from outside the field [59], it is critical to assess the state-of-the-art of EDA methodology at CHI. In the remainder of this paper, we review the literature discussed so far from a methodological perspective, assessing current practices and deriving recommendations for future work.

3 METHOD

The goal of this work is to assess and improve the quality of CHI studies involving the use of electrodermal activity data. Given the enormous potential of EDA data for revealing otherwise invisible aspects of the user experience, it is critical that this kind of data is collected, analysed, and reported appropriately. For this purpose, we conducted a systematic literature review of CHI papers involving EDA data as to how well they adhere to recommended practices from the field of psychophysiology.

3.1 Source selection

We reviewed full papers published in the proceedings of the ACM CHI Conference on Human Factors in Computing Systems, as representative of high-quality and mature research practices in the field of HCI—acknowledging that this excludes other venues interested in EDA data (e.g. Ubicomp, IMWUT, UIST, ISWC, etc). Most importantly, there are venues that publish HCI-related works that attract papers stemming from fields where EDA is more mature (e.g. medical, electrical engineering, signal processing), such as IEEE Sensors, IEEE Transactions on Biomedical Engineering, IEEE Transactions on Affective Computing, and IEEE Transactions on Cybernetics. As such, we restrict our claims to HCI work published at CHI. We conducted our search in the ACM digital library, selecting papers with the terms "Electrodermal Activity", "EDA", "Galvanic Skin Response", and "GSR" in any part of the paper. Our search spanned all CHI proceedings from 1986 to 2020. Specifically, we used the following search string:

```
{AllField:("eda" OR "electrodermal activity" OR
"galvanic skin response" OR "gsr")}
"filter": {Conference Collections: CHI: Conference
on Human Factors in Computing Systems}
```

3.2 Screening criteria

This search resulted in 235 papers. We further restricted this sample to only full research papers, resulting in 136 papers. The first author then inspected each paper in the sample and read the method section of each paper, removing those that did not include the collection of EDA data. The final sample was comprised of 41 papers, listed in the supplementary material.

3.3 Coding procedure

Our analysis is based on the *Publication recommendations for electrodermal measurements* devised by the Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures [13, 32]. The Society for Psychophysiological Research is a well-established society founded in 1960, which has commissioned recommendations for publication standards for EDA data since 1981, with the most recent edition of these standards published in 2012. As of the time of writing, the original set of recommendations had been cited over 1100 times and the most recent set of recommendations over 600 times, which is indicative of their widespread adoption. Other sources consulted in deriving our procedure included textbooks on EDA, such as Dawson et al. [26] and Boucsein [12].

Item	Description
Device	
1. Technology description	Technology used for recording EDA (Endosomatic, AC exosomatic, OR DC exomatic)
2. Applied voltage or current	If exomatic method used for recording, what is the applied current or voltage?
3. Device manufacturer and model	If a commercially available device is used, what is the name of the company and the model of the device?
4. Site description	Which body part is monitored (e.g. volar phalanges, thenar/hypothenar sites of the palm, foot)?
5. Sampling frequency	Rate at which the device captured data (measured in Hz)
6. Type and material of electrode	Type of electrode used (e.g. disposable, non-polarizing) and its material (e.g. carbon, Ag/AgCl)
7. Electrode fixation	How the electrodes were attached to the site to avoid motion artifacts (e.g. adhesive band tape)?
8. Contact area	Area of electrode skin contact (if possible to measure) and size of electrode (in cm^2 or diameter)
Pre-treatments and controls	
9. Pre-treatment on the skin	How the site was treated prior to attaching the electrodes (e.g. slight abrasion, lukewarm water, etc.)?
10. Electrode pre-treatment	Type of electrode pretreatment used (e.g. dry electrodes, pre-gelled disposable electrodes, or electrolyte)
11. Electrolyte	If electrolyte applied on electrodes, what is the type of gel and its concentration?
12. Pre-gelled disposable Electrodes	If pre-gelled disposable electrodes used, what is their brand and type?
13. Wait time before recording	After wearing a device, the experimenter must wait 5–10 minutes for participants to get a reasonable degree of hydration and then recording.
14. Duration of Recording	How long electrodes stayed in place?
15. Polarisation check	If using an endosomatic or DC exomatic EDA recording device, must control bias potential before start recording.
16. Polarisation reversal	If using DC exomatic EDA recording device (without non-polarizing electrodes), every 10–15 min (or between segments of each session) swap electrodes.
Participants and environment	
17. Age	Range, mean, and STD of participants' age is reported?
18. Sex/Gender	Sex/Gender of participants is reported?
19. Ethnicity	Ethnicity of participants is reported?
20. Caffeine	Was caffeine consumption controlled before the study?
21. Medication	Was medication consumption controlled before the study?
22. Temperature and Humidity	Temperature and humidity must be constant in all conditions of study
External factors	
23. Physical activity	Tasks in the study include intense physical activity?
24. Movement, speech, and deep breathing	Were participants asked to minimise movements, speech, and deep breathing?
25. Counter-balancing of external factors	Are external factors (e.g. movement, speech, and deep inhale) counter-balanced across conditions?
26. External factors within group	If comparing groups of participants, are attributes (gender, age, ethnicity) counter-balanced across groups?
Terminology and analysis	
27. Use of terminology	Are there any violations of standard terminology?
28. Smoothing EDA signal	Is any filtering or down-sampling reported?
29. Motion artefact removal	Is any pre-processing reported for correcting motion artifacts?
30. Range correction	Are any range correction or signal normalisation procedures for correcting individual differences reported?
31. Analysis method	Is the method of analysis (e.g. peak detection, model-based approach) described?
32. Parameterisation of raw values	Is any parameterisation of raw values (e.g. SCL, NS.SCR freq, SCR amplitude) described?
33. Model-based decomposition	If a model-based approach is used, are the details (name, software, any adjusted parameter in method) described?
34. Minimum amplitude for SCR and NS.SCR	If a peak detection approach is used to detect SCR and NS.SCR, are appropriate minimum amplitude criteria used and reported?
35. Time window for SCL	If a peak detection approach is used to measure SCL, are appropriate time window criteria used and reported?
36. Time window for SCR	If SCR is used for analysis, are appropriate delay window criteria used and reported?
37. Inter-Stimulus Interval (ISI)	If discrete stimuli were presented in the study, is an appropriate ISI reported?
38. Superimposed responses	If a peak detection approach is used to detect SCR and NS.SCR for analysis, how are superimposed responses treated?
Psychological constructs and parameters	
39. Psychological constructs	Is the physiological construct measurable by EDA?
40. Parameter	Are the parameters used for measuring the physiological construct, appropriately selected?

Table 2: Codebook used for the analysis of our sample

We coded each paper in terms of the information reported about the device(s) employed, any pre-treatment or controls, the participant sample and the environment where the study took place, the use of terminology, the study design, and the data analysis. Table 2 describes the information we captured for each item.

4 RESULTS

In this section, we describe the findings of our review, structured by the sections of our codebook. We describe problematic issues while highlighting examples of best practices found in this body of work.

4.1 Devices

Several different devices have been used in our sample (see Table 3). The most popular were the wearable sensors that emerged from the research of affective computing pioneer Rosalind Picard, manufactured by Empatica and Affectiva. As a reflection of the technical character of HCI work, many papers also employed custom-built devices, ranging from electrodes connected to low-cost hardware

Device	Number of papers used in
Empatica (E4, E3)	7
Custom-built device	4
NeXus-10 MKII	4
Shimmer 3 GSR	4
Biopac MP150	3
ProComp Infiniti system	3
BITalino (r)evolution Board Kit	3
Affectiva Q Sensor	2
Thought Technology's Triple Point Sensor	2
g.tec medical engineering	2
Flexcomp Infinity hardware	1
Varioport-B	1
Personal Input Pod	1
NeuroDyne MEDAC System/3	1
ADInstruments PowerLab	1
Grove GSR sensor device	1
H124SG Covidien	1
Microsoft Band	1

Table 3: EDA recording devices used by CHI papers

prototyping kits (e.g. Arduino [99] or Bitalino [79]) to novel EDA-sensing technologies (e.g. PrintSense [40]). Various medical-grade devices have also been employed.

Only one paper in our sample did not report which device they used to collect the EDA signal. However, with only two exceptions, most papers did not fully report the details of the sensor. Only seven papers explicitly reported the technology behind their sensor (Endosomatic/Exomatic DC/AC). Among the five of these that should have reported the details of the current or voltage, only three did so.

Seven papers made no mention of the recording site, while nine papers only partially described it. We coded as partial descriptions instances where the body part was not described in enough detail for precise replication. For example, only reporting that the device was attached to the "fingers", rather than specifying where on the finger—an example of a good site placement description can be found in Frison et al.: "*attached two skin electrodes to the volar (inner) middle phalanges (muscle limbs) of the non-dominant hand's middle and ring fingers*" [35]. Sixteen papers reported collecting data on the fingers, eleven on the wrist, three on the palm, and one on the shoulder. One paper used four sites simultaneously, namely the wrist, fingers, sole, and palm.

Fourteen papers failed to report the sampling rate of the device. The other ones that reported ranges between 1Hz and 512Hz, suggesting a large disparity in the granularity of the data collected.

There was generally very little information about electrodes in our sample. Only seven papers described the type of electrode, only five described the material, only six described the method of fixation of the electrodes, and only 4 measured the electrode size. The papers with the most comprehensive description of the electrodes described novel sensing devices [40, 68] rather than the use of a commercial device, though good examples can also be found in the latter category (e.g. [15, 85]).

4.2 Pre-treatments and controls

We identified severe under-reporting of pre-treatments required on the skin and electrodes, as well as of controls before and during the recording sessions. Only two papers described the pre-treatment they performed on the skin before placing the electrodes. As a good example of reporting pre-treatment on the skin we refer the reader to Putze et al. [78]: *"before placement of the electrodes, the participants cleaned their non-dominant hand with a wet wipe."*. Four papers reported the type of pre-treatment on the electrodes, three of which used dry electrodes (without any electrolytes, such as gels or pastes) which are prone to motion artifacts. Only one reported using electrolytes.

Because the electrolyte might affect the signal once it is applied, it is recommended to wait for 5–10 minutes after the electrodes are attached to begin the recording. It also allows the skin to have a suitable degree of hydration and minimises the impact of individual differences in sweating pattern prior to the study on EDA recording of the study [13]. Only four papers reported a wait time after wearing the EDA device to start recording. Two of them used an appropriate wait time (5 and 10 minutes), one used a very short wait time (1 minute), and one did not report this duration. None of the papers reported exactly how long the electrodes stayed in place. However, 27 of them reported the length of the study, which ranged from 7 minutes [89] to a whole day [63].

4.3 Participants and environments

Previous works have found that age [8, 17], gender [6, 17], and ethnicity [48, 56] affect the EDA signal, and as such, should be reported. For example, younger adults show larger increases in their SCR amplitude [38] and SCL [8] in cases of increasing arousal and males have a higher average SC than females [17]. According to Boucsein et al. [13], the number of participants, their gender, ethnicity, as well as the mean, range, and standard deviation of their age are the minimum required information that an EDA study should report about its participants to be comparable to other studies. One paper in our sample was excluded from this analysis as it did not describe the collection of EDA data [40].

None of the papers in our sample fully described their participant sample, though all of them did report the number of participants. Fifteen papers reported the necessary statistical information about ages (range, mean, and SD). The range of ages was not reported in 17 papers, the mean of ages was missing in 11 and the SD of ages was missing in 17. With the exception of 5 papers, most reported participants' gender distribution. Only one paper reported ethnicity, but only in general terms, stating that participants *"spanned varied different nationalities"*.

Caffeine [81] and medication [13] consumption, temperature[90] and relative humidity [7] have significant impacts on EDA signals, so they should be controlled for and reported. Eccrine sweat glands are influenced by temperature [13] so changes in temperature can affect EDA. Further, high humidity has also been shown to increase EDA levels [7]. Caffeine consumption induces arousal and increases EDA [24], while medications can alter EDA in different ways [36, 42]. However, based on our review, no CHI paper controlled humidity, only one reported the temperature, only one

controlled caffeine consumption and only three papers controlled medication consumption.

4.4 External factors

The impacts of external factors must be minimised and counter-balanced in different conditions of the study. Participant variables, movements, physical activities, speech, and deep inhalation are some of the external factors that could have a significant impact, and that should be controlled and reported in EDA studies.

Two studies in our sample compared groups of participants. Gender was counter-balanced in both and age was counter-balanced in one of them. They did not report any control for other factors and no normalisation for correcting individual differences. Sixteen papers in our sample included intense physical activities which may contain grasping, speech, and brisk movements, such as driving (4 papers), walking (5), nursing (2), playing an instrument (1), climbing (1), and running (1). Two papers recorded EDA in real workplaces where participants were not restricted in their actions. Seven papers informed participants about the artifacts that movements can cause and minimised it in their study. As a good example, see Frey et al. [34]: *"Participants were seated on a comfortable chair and asked to stay still throughout the recordings. [...] While the participant's non-dominant hand was resting on a flat surface to avoid EDA artifacts, their dominant hand was placed over a keyboard to advance between pages at their own pace."*

4.5 Terminology

In the background section, we described a standard terminology for referring to the EDA signal, and we evaluated how much of this terminology was adopted by our sample. Seven papers did not describe any EDA parameters whatsoever. Only 3 papers did not divert from the standard terminology. Eighteen papers used their own terminology, rather than the standard one (e.g. referring to the SCR amplitude as the "size of peak").

There were instances of papers trying to use the standard terminology, but with minor variations from it. Eight papers used tonic EDA terms for phasic EDA components and vice-versa. For instance, the term *"number of SCRs"* (a phasic term that refers to responses to a stimulus) was used in the analysis of tonic data without stimuli (where NS.SCR frequency is the appropriate term). Two papers used standard terms, but for the incorrect parameter. In one instance, the term SCR was used to refer to the amplitude of the peak of signal (the correct term should be "SCR amplitude"). Nine papers used the outdated acronym "GSR" when referring to EDA or its parameters.

4.6 Analysis

Raw EDA signals consist of tonic components, phasic components, noise, and artifacts due to electrode movements. Phasic and tonic EDA components should be parameterised and analysed separately [13]. To enable an informative analysis of the data, one must often down-sample the data, apply a low-pass filter, handle discontinuities in the signal, correct individual differences, and decompose the signal into its phasic and tonic components [22]. Each tonic and phasic component of EDA has specific parameters, which we described in Table 1. Meaningful analyses of EDA data are usually based on these parameters, which are chosen depending on the psychological construct of interest.

We did not consider three papers in this particular analysis. Two papers in our sample did not present the analysis of any EDA data, instead they focused on the description of novel EDA devices and used recorded EDA raw values to evaluate their proposed device. One paper used raw EDA signal values as the input for an artistic installation, and, as such, parameterisation was not relevant.

Most critically, of the remaining 38, 11 did not report any detail of how the signal was processed. Among the remaining 27 papers, 15 papers reported smoothing or down-sampling the signal in the pre-processing stage. Only 8 papers tried to correct motion artifacts in the signal and 14 papers performed range correction to minimise individual differences (this correction was not relevant in one of the papers, in which the authors compared simultaneous recordings from two devices attached to the same participant [89])

Of the 27 papers that detailed their analyses, 18 parameterised the signal, 7 presented no such parameterisation, and the information in the remaining 2 was not sufficient for us to determine whether the signal was parameterised. Among the 18 that parameterised the signal, 8 used peak detection, 4 used model-based approaches, and 6 used non-standard methods.

There are two main approaches for parameterising EDA signals—peak detection and model-based approaches. In the peak detection method, EDA parameters are extracted from SC signals by first identifying peaks within a time interval after the stimulus, using threshold criteria for a minimum rise in amplitude. Then, other parameters are extracted based on the position of the peak. The main challenge in peak detection methods is caused by super-imposed SCRs, in which an SCR starts before the previous one terminated. To address this problem, model-based approaches describe the sympathetic processes that generated the signal as a time series of a mathematical model. They then use the recorded EDA data to estimate the time-series process that generated the observed signal and derive the parameters from that.

Peak detection (8): When a peak detection method is used, the analysis should report the thresholding criteria and the time window over which it is computed. This is typically reported in terms of the minimum amplitude for identifying SCRs and NS.SCRs, the time window for measuring SCL, and the latency interval between the stimulus and the SCR [13]. In the 8 papers that used peak detection methods, 4 did not report the minimum amplitude. Among the remaining 4, three used thresholds within the recommended range of $0.01\mu S$ and $0.05\mu S$, while the remaining one used a value slightly above it ($0.07\mu S$). Because none of these papers used SCL, setting a time window was not relevant to them.

Model-based approaches (4): Four papers used model-based approaches. All of them specified the toolkit used, three of which used the MATLAB-based Ledalab [1] and one used python-based BioSPPy [2]. Only three specified the model, two of which used Continuous Decomposition Analysis (CDA) [9, 10] and one used cvxEDA [41].

Non-standard methods (6)/No parameterisation (7): Six papers used parameters other than the standard ones recommended

[1] http://www.ledalab.de/
[2] https://biosppy.readthedocs.io

in the literature. As a sample of non-standard parameterisation, papers used the average, standard deviation, minimum, and maximum of the signal as their parameters. Among these papers and the ones that did not perform any parameterisation, two papers reported averaging values within fixed time windows of 30s and one minute. Two papers reported using a moving average (3s and 5s) as a step in their signal smoothing process and one paper used a moving average window (3s) as a pre-processing step for measuring recovery time. However, due to the long duration of these windows, which is typically not suitable for smoothing, it was not clear whether this step was meant to remove noise or to remove the phasic component, isolating the SCL.

Handling super-imposed SCRs and SCR latency windows: As we previously discussed, super-imposed SCRs can create challenges for the analysis of EDA data. As such, it is important that papers report the inter-stimulus interval when using SCR-related parameters. Even though 7 papers used the phasic component of the signal in their analysis, only 2 reported the inter-stimulus interval, which in both cases was large enough to avoid super-imposed SCRs (1 min). Other than the papers that addressed super-imposed SCRs through the use of model-based approaches, of the 10 papers that used SCR or NS.SCR parameters, and as such, should have accounted for super-imposed SCRs, only two papers reported how they did so.

In order to ensure that an SCR relates to the intended stimulus, it is necessary to set a time window after the stimulus within which researchers can assume a causal relationship between the SCR and the stimulus. This window is typically set to be 1–4s after the stimulus. Five papers reported the SCR latency window, of which three used the standard 1–4s after the stimulus. One used 1–5s and one used 3–6s, which are outside of standard durations.

4.7 Psychological constructs

EDA reflects the activation of the sympathetic nervous system [26], which mediates involuntary responses to danger or stress [76]. This makes EDA a promising physiological signal to measure affective and cognitive states [22]. In our sample, different psychological constructs were operationalised through EDA, namely arousal, stress, orienting response, cognitive load (mental workload), valence, emotion (affect, emotional reaction), flow, and enjoyment. Previous work has shown an association between EDA and some of these constructs, namely arousal, stress, orienting response, and cognitive load (mental workload) (27 papers in our sample).

In contrast, psychological constructs such as flow, enjoyment, valence, and engagement mostly refer to positive mental states. In previous work, EDA is mostly considered a marker of negative cognitive activity [12], showing no significant relationship with positive mental states such as valence [83]. We could not find evidence to support the physiological constructs used in 5 papers in our sample.

Arousal: Arousal is the most common construct measured with EDA in psychology. Niven and Miles [69] define arousal as a *"state of feeling awake, activated, and highly reactive to stimuli"*. For Boucsein [12], EDA is an indicator of lower levels of arousal, which is mostly caused by cognitive processes. The psychology literature supports the use of the tonic EDA signal as an indicator of arousal (e.g. Dawson et al. [26]). SCL and NS.SCR frequency are tonic measures widely used to measure arousal [13].

Examples of EDA being used to measure arousal abound in the CHI literature. Fifteen papers in our sample reported using EDA to measure or classify arousal. Only two of them, however, used parameters aligned with the literature: Frison et al. [35] used the *"number of SCRs"* (NS.SCR frequency) as an indicator of arousal of participants using a driving simulator; and Mandryk et al. [58] used SCL to measure arousal in the context of the Arousal-Valence model of emotion. Two other papers that reported arousal classification using EDA features used a combination of valid and invalid features. McDuff et al. [63] used different features extracted from EDA to classify arousal. However, except for one, *"number of peaks"* (NS.SCR frequency), the features were not standard EDA parameters with evidence in the literature as measures of arousal. Frey et al. [34] used three EDA features to classify arousal while reading a story. They used *"total number of peaks"* (NS.SCR frequency), *"average amplitude"* of peaks, and average Sudometer Nerve Activity (a feature computed by the cvxEDA package). Except for the first one, we found no support in the literature for the other two.

Two other papers reported using SCL to measure arousal, but with issues in the computation of the SCL. Matthews et al. [61] used SCL as an indicator of arousal in a biofeedback system. Although their parameter choice for measuring EDA is correct, they used Least Mean Square of EDA in successive windows, which is not a standard way of measuring SCL. Mirza-Babaei et al. [66] reported using the SCL to measure players' arousal, but the lack of detail in their reporting suggests they used raw values.

Nine papers used inappropriate parameters as measures of arousal. Four of them used phasic parameters, such as the NS.SCR amplitude. NS.SCRs themselves are not related to arousal, but rather their frequency is the parameter that the literature recommends using. In one case, phasic parameters were used to measure arousal. However, the SCR parameters they used are commonly associated with novelty or intensity of reactions to a stimulus—not arousal.

Five of the papers that used an inappropriate measure for arousal employed EDA values without performing any parameterisation. These included raw values, the difference between the maximum and minimum values of the signal, baseline-adjusted values, and normalised values. There is no support in the literature for any of these as measures of arousal.

Stress: EDA is controlled by the sympathetic nervous system, which is the branch of the nervous system predominantly responsible for feelings of stress. This makes EDA a very informative signal for determining when a user is experiencing stress [12]. Boucsein [12] suggests the NS.SCR frequency as a valid measure of stress.

Five papers in our sample measured stress using EDA, but only two used appropriate parameters. Schulz et al. [82] used the *"conductivity level"* (SCL) and *"response count"* (NS.SCR frequency), while Paredes et al. [74] used the *"number of phasic peaks"* (NS.SCR frequency).

One paper investigated the correlation between stress and the *"number of peaks"* (NS.SCR frequency), the *"average EDA"* and the *"range of values"*. Among these, the only parameter with support in the literature for measuring stress is the NS.SCR frequency.

One paper used EDA alongside heart rate and respiration rate to detect stress but did not report which parameter was used to

operationalise it. Another paper used EDA alongside blood pressure and respiration rate to detect stress in video mediated collaboration. However, they did not report any information about EDA preprocessing and parameters they used. A final one used normalised raw values to measure the correlation between stress and EDA recorded from two different devices (Microsoft band and Nexus 10).

Orienting Response: Phasic SCRs are reliable indicators of *orienting responses*. An orienting response is the physiological reaction of the body to a stimulus or to changes in a stimulus [84]. In this type of study, researchers monitor phasic SCRs while participants observe a stimulus to study the impact of that stimulus on the participants. If participants exhibit SCRs after perceiving the stimulus, it suggests that the stimulus is novel to them. The intensity of a stimulus can then be measured using the SCR amplitude—more intense stimuli cause higher SCR amplitudes. In HCI, orienting responses have applications in measuring the impacts of short-lasting events, such as notifications or interruptions, on the user experience. For instance, if an SCR is observed after notification, it is likely that the user perceived it and it was novel to them [31]. The higher the amplitude of an SCR, the higher the intensity of this notification effect on the user.

Five papers in our sample used EDA to measure novelty or intensity of discrete stimuli on the user experience, three of which employed appropriate parameters. Ogawa et al. [70] used the SCR amplitude to measure the significance of different threats (e.g. virtual needles popping out from a virtual wall) in a VR environment expecting participants with greater sense of presence shows higher SCRs in response to the threat. Fortin et al. [31] measured the *"maximum of the phasic activity"* (SCR amplitude) when users received a notification to see whether the perception of a notification causes SCRs. Pan et al. [73] used the SCR amplitude to detect external interruptions while listening to an audio recording.

One of the papers, used a combination of phasic (*"maximum peak amplitude"*, *"number of peaks"*) and tonic measures (*"mean SCL"*) to evaluate the impact of different driving conditions (e.g. stopping behind a parked car, passing a parked car, etc.) on drivers. Because the stimuli in this study were fast-changing, slow-changing tonic parameters are usually not recommended. Another paper that investigated the effects of interruptions in the user experience computed the difference between values within a 3–6s window after the interruption to a baseline. In addition, they computed the difference between the SCL (in 3s moving average windows) to the baseline to measure the recovery time. More appropriate measures of the novelty and intensity of a discrete stimulus such as an interruption would be phasic parameters. More appropriate measures for measuring the recovery time would be the standard rec.tc and rec.t/2 parameters.

Cognitive Load: Cognitive load is defined as a *"multidimensional construct that represents the load that performing a particular task imposes on the cognitive system of a learner"*[72]. SCR is considered a reliable index of cognitive resource demand [52]. However, even though there is some support in previous works for measuring cognitive load with EDA [19, 46, 83, 94], CHI authors should use EDA cautiously for this purpose, because other works have found conflicting results [44, 44, 54]. Charles and Nixon [18] reviewed papers that used EDA to measure cognitive load and concluded that there has not been enough evidence indicating that EDA can differentiate between lower and higher mental workloads. However, it can indicate mental workload differences between different task types and increases in mental workload demand.

Cognitive load was measured in three papers in our sample using EDA. One of the papers did not use an appropriate parameter for measuring it. This paper used the *"mean of the non-negative first-order differential of EDA signals"* in different study conditions to measure significant changes in mental workload, but we were not able to map this feature to any standard parameter. Another used the mean, STD, minimum, maximum, and first derivative of the SCL in a model to classify the cognitive load of drivers. Though SCL can be representative of cognitively induced changes in EDA, this approach is not standard. One paper used EDA to measure the cognitive load of participants working with different parts of a website, but did not report how the authors derived cognitive load from the EDA signal.

5 DISCUSSION

So far, we reviewed the CHI literature in terms of current practices related to electrodermal activity data. Through a review of 41 CHI papers, we identified challenges in the way that these data are collected, analysed, and reported. In this section, we highlight the most critical issues and discuss potential consequences if they are left unaddressed. These include threats to internal and external validity, study replicability, and general best practices. We acknowledge the challenges we described were identified in the CHI literature and analysing EDA practices in other HCI venues such as Ubicomp, IMWUT, UIST, and ISWC may lead to different outcomes.

5.1 Threats to Validity

Compromising internal validity is among the most serious issues as any deductions made from studies that fall victim to it are potentially confounded. Threats to external validity, on the other hand, are often factors that refer to the specific study conditions rather than the general nature of EDA. Based on our review, it was not always clear whether internal validity had been preserved as procedure descriptions were often limited. For example, considering how devices were worn. Only a few papers reported the fixation method for the electrode (e.g. tape) and even fewer the use of electrolytes. Pre-treatments on the skin (e.g. washing with lukewarm water) and waiting for a suitable degree of hydration were mostly not considered. We observed that the same situation prevailed in how controls were performed before and during the studies. Most of the papers did not acknowledge or control for the impacts of polarisation on the recordings, and none of them swapped electrodes to minimise polarisation bias. Temperature, humidity, medication, and caffeine consumption were controlled in relatively few papers.

Another important concern is the site on the skin from where the data is collected in many cases. Sweating patterns in different parts of the skin can be controlled by various internal processes. Therefore, EDA recordings are only valid when electrodes are placed on a part of the skin in which sweating is dominantly controlled by the sympathetic nervous system as opposed to thermoregulatory processes. The most widely recommended sites for placing electrodes are the distal phalanges on the non-dominant hand [33]. However, this site is not always available based on study conditions

or activities. As alternatives, the medial phalanges, thenar, and hypothenar sites on the hand and the inner aspect of the feet can also be used for recording EDA data. Several papers in our sample used sites on the skin (e.g wrist and shoulder) of which changes in conductivity are not representative of sympathetic nervous system processes. The wrist is a very popular site for EDA recording in CHI papers; however, there is some evidence in the literature that sweat gland activity on the skin of the wrist is mostly representative of thermoregulatory activity, rather than EDA [13] and several validation studies reported unreliability of EDA recordings from wrist measurements [64, 65, 71, 75, 92, 93].

Physical activities can affect EDA recordings in two ways: destabilising electrodes and increasing body heat. EDA recordings depend on the skin-electrode contact area [57], and even small movements of an unstable electrode may change its contact area with the skin, creating artifacts in the recorded signal. This issue is even more severe in CHI studies, as our review showed that most of the papers used dry electrodes (electrodes without electrolytes) which become unstable after hydration of the skin [13]. Physical activities in EDA studies may lead to movements of the electrodes, thereby affecting the signal. Further, intense physical activities increase the body temperature and consequently, cause increased sweating to reduce it, which also affects the EDA signal. Processing the EDA signal in this situation requires both qualitative and quantitative information about the physical activities involved, which is rarely recorded or accounted for by modern devices [13]. Though typical studies involving EDA try to minimise physical activity as much as possible, many works at CHI involve strenuous tasks, from sprint cycling [5] to climbing [82], which put the quality of the EDA data at risk.

Speech, social interactions, deep inhalation or even different postures such as standing or sitting are other extraneous factors that may impact the signal. Therefore, it is important to control for differences in these factors in the different study conditions. For example, in a study in our sample, only one of the conditions included physical activities, confounding the effect of the variable of interest with the effect of the physical activity.

5.2 Replicability

Echoing recent concerns regarding the lack of transparency in statistical reporting in HCI [21, 47], we note the lack of clarity and detail required for replicating studies involving the use of EDA data. In several cases, descriptions and methods were not comprehensible due to selective, incomplete, and non-standard presentation of information. In most papers, we could not find information on EDA devices, participants' details, procedures, signal processing, and analysis. Violation of standard terminology, using self-defined terms for EDA parameters, and confusion between phasic and tonic parameters made understanding papers cumbersome and in some cases impossible for the reader. Not adopting the standard terminology hinders the clarity of these papers, but more importantly, it threatens the community's ability to retrace and replicate these studies. In this paper, we clarified the terminology around EDA and what is required to properly plan, conduct, and report studies with it.

5.3 Adhering to Best Practices

Following best practices reduces the risk of many of the issues mentioned above. Further, it produces better datasets as noise in measurements can be effectively reduced. To ensure high-quality EDA data, for example, different treatments must be applied on the skin or electrodes prior to the recording, and several external factors must be controlled during the recording. In ambulatory recordings used for in-situ studies, some of these steps may be skipped due to the limited control on the participant and their environment. However, laboratory studies should adhere to all of these steps as much as possible. Although the majority of studies in our sample (32) were conducted in a lab setting, there is an ongoing trend towards ambulatory recordings. This affects the choice of pre-treatments, devices, and controls.

Given the drive to evaluate technologies in realistic environments, many of the devices employed are suitable for using ambulatory recording devices, such as the Shimmer 3 GSR and the Empatica E4. However, although these devices offer more mobility for participants, they are not as accurate as conventional devices [45, 89, 93]. As such, unless the experiment requires participant mobility, stationary recording devices should be used instead.

Despite the importance of data pre-processing and signal parameterisation in the analysis of EDA data, we identified a concerning pattern among the papers in our sample of using raw values directly. Raw EDA values are an aggregation of tonic and phasic EDA components, noise, motion artifacts, and the effect of the variables of interest. The meaningful parts of the data are the phasic and tonic EDA components, so it is necessary to isolate them before analysing the data. The fact that 11 papers did not report any details on neither the analysis nor the data processing is suggestive of the use of raw values. Among the rest of the papers, one or more pre-processing procedures were skipped, and only 12 papers used standard methods for processing the EDA signal. This suggests that the CHI community should improve its adoption of standards and well-established methods for EDA analysis.

5.4 Limitations

As a sample of HCI research involving EDA, we selected 41 papers from the ACM CHI Proceedings. However, HCI covers a very broad range of proceeding and journals. Therefore, the fact that our selection of papers may not be representative of all types of EDA research in HCI is the first limitation of this review.

Further, the lack of guidelines for EDA research in HCI and associated fields led us to draw from the psychophysiology literature to extract guidelines to evaluate our sample. We acknowledge that the objectives of EDA research in psychophysiology for which these guidelines were developed may be somewhat different from the objectives of EDA research in HCI. This may potentially make our review overly strict when evaluating certain practices. As an example of this issue, we can refer to the few papers in our sample that used EDA for artistic purposes, which passed the inclusion criteria even though their objectives and methods are rather different to those in psychophysiology. Nevertheless, we argue that even an artist interested in incorporating physiological signals in their installations would require the signals that their sensors are capturing to reflect the actual psychological processes that inspired

the artwork in the first place—if this did not matter, there are more convenient sources of noise that they could have used instead.

Having said that, a significant proportion of our sample dealt with phenomena that are also objects of study in psychophysiology (e.g. arousal, orienting responses, cognitive load), so the advice from that community is nevertheless valid. Moreover, considering the broad range of HCI research, coming up with guidelines and framework that fit every area of HCI research is impractical, if not impossible.

6 COMMUNITY-SOURCED GUIDELINES

Given the variety of uses of EDA in HCI, we believe that the best way to improve community standards is to co-design such standards given the particular needs and practices of HCI researchers. As such, we take inspiration from the Special Interest Group on Transparent Statistics in HCI, who have been running workshops at CHI since 2016 [50, 51, 97] and have since created a wiki for the HCI community to contribute towards better statistical practices [47]. We have developed an initial set of guidelines published at https://edaguidelines.github.io/, which we invite the community to contribute their standards and best practices, according to the various needs that arise in HCI practice and research. The developed initial set of guidelines are structured based on steps of an EDA study (e.g. design, recruiting participants, data collection, etc.) and focus on issues with the highest concern in the reviewed CHI samples. These recommendations aim at assuring a required minimum scientific quality for future CHI EDA practices.

For a deeper comprehension and a more fundamental knowledge of EDA methods, researchers should refer to standard texts, such as the *Publication recommendations for electrodermal measurements* devised by the Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures [13], *The Electrodermal System* [26] chapter in the *Handbook of Psychophysiology* [4], and the book *Electrodermal Activity* by Boucsein [12].

7 RESEARCH GAPS AND FUTURE WORK

Though our review identified problematic practices in the use of EDA data by the CHI community, there are several directions in which CHI researchers can make novel methodological contributions to this field.

Devices for ambulatory recording: Through its drive to develop and deploy technologies in realistic settings to address the needs of real users, CHI researchers are well-equipped to make solid contributions to the design and evaluation of ambulatory EDA recording devices. Though technologies for recording EDA data in the lab are mature, new technologies for recording EDA in ambulatory setting are urgently required [26]. The CHI literature already contains examples of novel EDA recording devices [40, 68]; but more research is necessary on improved ambulatory recording devices that are robust to the lack of control in everyday settings.

Validation of existing devices: Different devices have been used in CHI papers presented in Table 3; however, more work on the validation of these recording devices in the contexts of interest to CHI is urgently needed. This is particularly evidenced by the fact that previous works have reported inaccurate recordings [89] and inconsistent results [45] from some of these devices, which renderss the need for validation studies at CHI even more critical.

User-friendly interactive toolkits: Although toolkits for EDA pre-processing and analysis are available [1–3], most of them are implemented in technical packages with a steep learning curve. Several of them are implemented in MATLAB, which not only requires that users have experience with this platform but also makes it difficult to incorporate the analyses into real-time, interactive systems. Therefore, novel user-friendly signal processing toolkits for real-time or interactive analysis of EDA data would benefit the HCI community both by improving the validity of studies involving EDA data and by opening new opportunities for better uses of EDA as an input modality for interactive systems.

Replications of EDA Research: Cockburn recently warned the community that a replication crisis is looming upon us [20]. Considering the methodological issues in the existing body of EDA research at CHI, the necessity for replications of EDA studies is evident. The replication of previous EDA studies gives us the chance to reexamine the conclusions from these studies and improve the confidence in the results reported so far. Such replication efforts should take into account the recommendations laid out in this paper to ensure that they do not fall into the same traps as previous works.

8 CONCLUSION

In this paper, we presented a methodological review of CHI papers that involve the use of electrodermal activity data. We found severe methodological issues in this body of work, ranging from the collection, over analysis, to the reporting of EDA data. Problematic areas included a lack of transparency and inadequate reporting of procedures and results, the incorrect use of recording equipment, poor controls in study designs, and the misuse of signal processing techniques and analysis procedures. Our findings suggest an urgent need for improved practices and increased awareness of methodological standards.

ACKNOWLEDGMENTS

Dr Eduardo Velloso is the recipient of an Australian Research Council Discovery Early Career Award (Project Number: DE180100315) funded by the Australian Government and Ebrahim Babaei is recipient of the Melbourne Research Scholarship.

REFERENCES

[1] [n.d.]. API Reference — BioSPPy 0.6.1 documentation. https://biosppy.readthedocs.io/en/stable/biosppy.html. (Accessed on 09/17/2020).
[2] [n.d.]. Ledalab. http://www.ledalab.de/. (Accessed on 09/17/2020).
[3] [n.d.]. PsPM | A matlab suite for Psycho-Physiological Modelling. http://pspm.sourceforge.net/. (Accessed on 09/17/2020).
[4] 2016. *Handbook of Psychophysiology* (4 ed.). Cambridge University Press. https://doi.org/10.1017/9781107415782
[5] Soumya C. Barathi, Michael Proulx, Eamonn O'Neill, and Christof Lutteroth. 2020. Affect Recognition Using Psychophysiological Correlates in High Intensity VR Exergaming. In *Proc. of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376596
[6] Dindar Bari. 2020. Gender Differences in Tonic and Phasic Electrodermal Activity Components. *Science Journal of University of Zakho* 8, 1 (Mar. 2020), 29–33. https://doi.org/10.25271/sjuoz.2020.8.1.670
[7] Dindar S. Bari, Haval Y. Yacoob Aldosky, Christian Tronstad, Håvard Kalvøy, and Ørjan G. Martinsen. 2018. Influence of Relative Humidity on Electrodermal Levels and Responses. *Skin Pharmacology and Physiology* 31, 6 (2018), 298–307. https://doi.org/10.1159/000492275

[8] Dindar S. Bari, Haval Y. Yacoob Aldosky, and Ørjan G. Martinsen. 2020. Simultaneous measurement of electrodermal activity components correlated with age-related differences. *Journal of Biological Physics* 46, 2 (01 Jun 2020), 177–188. https://doi.org/10.1007/s10867-020-09547-4

[9] Mathias Benedek and Christian Kaernbach. 2010. A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods* 190, 1 (2010), 80 – 91. https://doi.org/10.1016/j.jneumeth.2010.04.028

[10] Mathias Benedek and Christian Kaernbach. 2010. Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology* 47, 4 (2010), 647–658. https://doi.org/10.1111/j.1469-8986.2009.00972.x

[11] Andrew Best and Jason M. Kamilar. 2018. The evolution of eccrine sweat glands in human and nonhuman primates. *Journal of Human Evolution* 117 (2018), 33 – 43. https://doi.org/10.1016/j.jhevol.2017.12.003

[12] Wolfram Boucsein. 2012. *Electrodermal Activity.* Springer US. https://doi.org/10.1007/978-1-4614-1126-0

[13] Wolfram Boucsein, Don C. Fowles, Sverre Grimnes, Dgershon Ben-Shakhar, Walton T. Roth, Michael E. Dawson, and Diane L. Filion. 2012. Publication recommendations for electrodermal measurements. *Psychophysiology* 49, 8 (2012), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

[14] Clinton C. Brown. 1967. A proposed standard nomenclature for psychophysiologic measures. *Psychophysiology* 4, 2 (1967), 260–264. https://doi.org/10.1111/j.1469-8986.1967.tb02765.x

[15] Anders Bruun, Effie Lai-Chong Law, Matthias Heintz, and Lana H.A. Alkly. 2016. Understanding the Relationship between Frustration and the Severity of Usability Problems: What Can Psychophysiological Data (Not) Tell Us?. In *Proc. of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 3975–3987. https://doi.org/10.1145/2858036.2858511

[16] Delphine Caruelle, Anders Gustafsson, Poja Shams, and Line Lervik-Olsen. 2019. The use of electrodermal activity (EDA) measurement to understand consumer emotions – A literature review and a call for action. *Journal of Business Research* 104 (2019), 146 – 160. https://doi.org/10.1016/j.jbusres.2019.06.041

[17] Steve Chamberlin, Agatha P. Colbert, and Adrian Larsen. 2011. Skin Conductance at 24 Source (Yuan) Acupoints in 8637 Patients: Influence of Age, Gender and Time of Day. *Journal of Acupuncture and Meridian Studies* 4, 1 (2011), 14 – 23. https://doi.org/10.1016/S2005-2901(11)60003-4

[18] Rebecca L. Charles and Jim Nixon. 2019. Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics* 74 (2019), 221 – 232. https://doi.org/10.1016/j.apergo.2018.08.028

[19] Fang Chen, Jianlong Zhou, Yang Wang, Kun Yu, Syed Z. Arshad, Ahmad Khawaji, and Dan Conway. 2016. *Robust Multimodal Cognitive Load Measurement.* Springer International Publishing. https://doi.org/10.1007/978-3-319-31700-7

[20] Andy Cockburn, Pierre Dragicevic, Lonni Besançon, and Carl Gutwin. 2020. Threats of a Replication Crisis in Empirical Computer Science. *Commun. ACM* 63, 8 (July 2020), 70–79. https://doi.org/10.1145/3360311

[21] Andy Cockburn, Carl Gutwin, and Alan Dix. 2018. HARK No More: On the Preregistration of CHI Experiments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173715

[22] Benjamin Cowley, Marco Filetti, Kristian Lukander, Jari Torniainen, Andreas Henelius, Lauri Ahonen, Oswald Barral, Ilkka Kosunen, Teppo Valtonen, Minna Huotilainen, Niklas Ravaja, and Giulio Jacucci. 2016. The Psychophysiology Primer: A Guide to Methods and a Broad Review with a Focus on Human–Computer Interaction. *Foundations and Trends® in Human–Computer Interaction* 9, 3-4 (2016), 151–308. https://doi.org/10.1561/1100000065

[23] Hugo D. Critchley. 2002. Electrodermal responses: What happens in the brain. *Neuroscientist* 8, 2 (2002), 132–142. https://doi.org/10.1177/107385840200800209

[24] Robyn A. Davidson and Barry D. Smith. 1991. Caffeine and novelty: Effects on electrodermal activity and performance. *Physiology & Behavior* 49, 6 (1991), 1169 – 1175. https://doi.org/10.1016/0031-9384(91)90346-P

[25] Michael E. Dawson, Anne M. Schell, and Diane L. Filion. 2007. *The electrodermal system.* Cambridge University Press, New York, NY, US, 159–181. https://doi.org/10.1017/CBO9780511546396.007

[26] Michael E. Dawson, Anne M. Schell, and Diane L. Filion. 2017. *The electrodermal system.* Cambridge University Press, New York, NY, US, 217–243.

[27] Nicole Dillen, Marko Ilievski, Edith Law, Lennart E. Nacke, Krzysztof Czarnecki, and Oliver Schneider. 2020. Keep Calm and Ride Along: Passenger Comfort and Anxiety as Physiological Responses to Autonomous Driving Styles. In *Proc. of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376247

[28] Vanessa Echeverria, Roberto Martinez-Maldonado, and Simon Buckingham Shum. 2019. Towards Collaboration Translucence: Giving Meaning to Multimodal Group Data. In *Proc. of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, 1–16. https://doi.org/10.1145/3290605.3300269

[29] Peter H. Ellaway, Annapoorna Kuppuswamy, Alessia Nicotra, and Christopher J. Mathias. 2010. Sweat production and the sympathetic skin response: Improving the clinical assessment of autonomic function. *Autonomic Neuroscience: Basic and Clinical* 155, 1-2 (2010), 109–114. https://doi.org/10.1016/j.autneu.2010.01.008

[30] Lara Fontanella, Luigi Ippoliti, and Arcangelo Merla. 2012. Multiresolution Karhunen Loéve analysis of galvanic skin response for psycho-physiological studies. *Metrika* 75, 3 (2012), 287–309. https://doi.org/10.1007/s00184-010-0327-3

[31] Pascal E. Fortin, Elisabeth Sulmont, and Jeremy Cooperstock. 2019. Detecting Perception of Smartphone Notifications Using Skin Conductance Responses. In *Proc. of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, 1–9. https://doi.org/10.1145/3290605.3300420

[32] Don C. Fowles, Margaret J. Christie, Robert Edelberg, William W. GRINGS, David T. Lykken, and Peter H. Venables. 1981. Publication Recommendations for Electrodermal Measurements. *Psychophysiology* 18, 3 (1981), 232–239. https://doi.org/10.1111/j.1469-8986.1981.tb03024.x

[33] Lauren W. Freedman, Angela Scarpa Scerbo, Michael E. Dawson, Adrian Raine, William O. Mcclure, and Peter H. Venables. 1994. The relationship of sweat gland count to electrodermal activity. *Psychophysiology* 31, 2 (1994), 196–200. https://doi.org/10.1111/j.1469-8986.1994.tb01040.x

[34] Jérémy Frey, Gilad Ostrin, May Grabli, and Jessica R. Cauchard. 2020. Physiologically Driven Storytelling: Concept and Software Tool. In *Proc. of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376643

[35] Anna-Katharina Frison, Philipp Wintersberger, Andreas Riener, Clemens Schartmüller, Linda Ng Boyle, Erika Miller, and Klemens Weigl. 2019. In UX We Trust: Investigation of Aesthetics and Usability of Driver-Vehicle Interfaces and Their Impact on the Perception of Automated Driving. In *Proc. of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300374

[36] C. D. Frith, Marilyn Stevens, Eve C. Johnstone, and D. G. C. Owens. 1984. The effects of chronic treatment with amitriptyline and diazepam on electrodermal activity in neurotic outpatients. *Physiological Psychology* 12, 3 (Sept. 1984), 247–252. https://doi.org/10.3758/bf03332199

[37] Wei Gai, Chenglei Yang, Yulong Bian, Chia Shen, Xiangxu Meng, Lu Wang, Juan Liu, Mingda Dong, Chengjie Niu, and Cheng Lin. 2017. Supporting Easy Physical-to-Virtual Creation of Mobile VR Maze Games: A New Genre. In *Proc. of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. ACM, New York, NY, USA, 5016–5028. https://doi.org/10.1145/3025453.3025494

[38] Joachim Gavazzeni, Stefan Wiens, and Håkan Fischer. 2008. Age effects to negative arousal differ for self-report and electrodermal activity. *Psychophysiology* 45, 1 (2008), 148–151. https://doi.org/10.1111/j.1469-8986.2007.00596.x

[39] Vanessa Georges, François Courtemanche, Sylvain Senecal, Thierry Baccino, Marc Fredette, and Pierre-Majorique Leger. 2016. UX Heatmaps: Mapping User Experience on Visual Interfaces. In *Proc. of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 4850–4860. https://doi.org/10.1145/2858036.2858271

[40] Nan-Wei Gong, Jürgen Steimle, Simon Olberding, Steve Hodges, Nicholas Edward Gillian, Yoshihiro Kawahara, and Joseph A. Paradiso. 2014. PrintSense: A Versatile Sensing Technique to Support Multimodal Flexible Surface Interaction. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 1407–1410. https://doi.org/10.1145/2556288.2557173

[41] Alberto Greco, Gaetano Valenza, Antonio Lanata, Enzo Scilingo, and Luca Citi. 2016. cvxEDA: a Convex Optimization Approach to Electrodermal Activity Processing. *IEEE Transactions on Biomedical Engineering* (2016), 1–1. https://doi.org/10.1109/tbme.2015.2474131

[42] Michael Foster Green, Keith H. Nuechterlein, and Paul Satz. 1989. The Relationship of Symptomatology and Medication to Electrodermal Activity in Schizophrenia. *Psychophysiology* 26, 2 (1989), 148–157. https://doi.org/10.1111/j.1469-8986.1989.tb03147.x

[43] Diane Gromala, Xin Tong, Amber Choo, Mehdi Karamnejad, and Chris D. Shaw. 2015. The Virtual Meditative Walk: Virtual Reality Therapy for Chronic Pain Management. In *Proc. of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 521–524. https://doi.org/10.1145/2702123.2702344

[44] Eija Haapalainen, SeungJun Kim, Jodi F. Forlizzi, and Anind K. Dey. 2010. Psycho-Physiological Measures for Assessing Cognitive Load. In *Proc. of the 12th ACM International Conference on Ubiquitous Computing* (Copenhagen, Denmark) *(UbiComp '10)*. ACM, New York, NY, USA, 301–310. https://doi.org/10.1145/1864349.1864395

[45] Katrin Hänsel, Romina Poguntke, Hamed Haddadi, Akram Alomainy, and Albrecht Schmidt. 2018. What to Put on the User: Sensing Technologies for Studies and Physiology Aware Systems. In *Proc. of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173719

[46] C. S. Ikehara and M. E. Crosby. 2005. Assessing Cognitive Load with Physiological Sensors. In *Proceedings of the 38th Annual Hawaii International Conference on*

System Sciences. 295a–295a. https://doi.org/10.1109/HICSS.2005.103

[47] Transparent Statistics in Human–Computer Interaction Working Group. 2019. Transparent Statistics Guidelines. https://doi.org/10.5281/zenodo.1186169 (Available at https://transparentstats.github.io/guidelines).

[48] Laverne C. Johnson and Marvin M. Landon. 1965. Eccrine sweat gland activity and racial differences in resting skin conductance. *Psychophysiology* 1, 4 (1965), 322–329. https://doi.org/10.1111/j.1469-8986.1965.tb03264.x

[49] Cornelia Kappeler-Setz, Franz Gravenhorst, Johannes Schumm, Bert Arnrich, and Gerhard Tröster. 2013. Towards long term monitoring of electrodermal activity in daily life. *Personal and Ubiquitous Computing* 17, 2 (01 Feb 2013), 261–271. https://doi.org/10.1007/s00779-011-0463-4

[50] Matthew Kay, Steve Haroz, Shion Guha, and Pierre Dragicevic. 2016. Special Interest Group on Transparent Statistics in HCI. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 1081–1084. https://doi.org/10.1145/2851581.2886442

[51] Matthew Kay, Steve Haroz, Shion Guha, Pierre Dragicevic, and Chat Wacharamanotham. 2017. Moving Transparent Statistics Forward at CHI. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 534–541. https://doi.org/10.1145/3027063.3027084

[52] Arthur E Kramer. 1990. Physiological Metrics of Mental Workload: A Review of Recent Progress.

[53] Kai Kuikkaniemi, Toni Laitinen, Marko Turpeinen, Timo Saari, Ilkka Kosunen, and Niklas Ravaja. 2010. The Influence of Implicit and Explicit Biofeedback in First-Person Shooter Games. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. ACM, New York, NY, USA, 859–868. https://doi.org/10.1145/1753326.1753453

[54] Charlotte Larmuseau, Pieter Vanneste, Jan Cornelis, Piet Desmet, and Fien Depaepe. 2019. Combining physiological data and subjective measurements to investigate cognitive load during complex learning. *Frontline Learning Research* 7, 2 (2019), 57–74. Uhttps://lirias.kuleuven.be/retrieve/539018DLarmuseauetal.-3825-2-10-20190510.pdf[freelyavailable]

[55] Celine Latulipe, Erin A. Carroll, and Danielle Lottridge. 2011. Love, Hate, Arousal and Engagement: Exploring Audience Responses to Performing Arts. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 1845–1854. https://doi.org/10.1145/1978942.1979210

[56] Israel Lieblich, Sol Kugelmass, and Gershon Ben-Shakhar. 1973. Psychophysiological Baselines as a Function of Race and Ethnic Origin. *Psychophysiology* 10, 4 (1973), 426–430. https://doi.org/10.1111/j.1469-8986.1973.tb00800.x

[57] Mary L. Mahon and William G. Iacono. 1987. Another Look at the Relationship of Electrodermal Activity to Electrode Contact Area. *Psychophysiology* 24, 2 (1987), 216–222. https://doi.org/10.1111/j.1469-8986.1987.tb00280.x

[58] Regan L. Mandryk, M. Stella Atkins, and Kori M. Inkpen. 2006. A Continuous and Objective Evaluation of Emotional Experience with Interactive Play Environments. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI '06)*. ACM, New York, NY, USA, 1027–1036. https://doi.org/10.1145/1124772.1124926

[59] Joe Marshall and Conor Linehan. 2017. Misrepresentation of Health Research in Exertion Games Literature. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 4899–4910. https://doi.org/10.1145/3025453.3025691

[60] Roberto Martinez-Maldonado, Vanessa Echeverria, Gloria Fernandez Nieto, and Simon Buckingham Shum. 2020. From Data to Insights: A Layered Storytelling Approach for Multimodal Learning Analytics. In *Proc. of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376148

[61] Mark Matthews, Jaime Snyder, Lindsay Reynolds, Jacqueline T. Chien, Adam Shih, Jonathan W. Lee, and Geri Gay. 2015. Real-Time Representation Versus Response Elicitation in Biosensor Data. In *Proc. of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 605–608. https://doi.org/10.1145/2702123.2702485

[62] Jon McCormack, Toby Gifford, Patrick Hutchings, Maria Teresa Llano Rodriguez, Matthew Yee-King, and Mark d'Inverno. 2019. In a Silent Way: Communication Between AI and Improvising Musicians Beyond Sound. In *Proc. of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, 1–11. https://doi.org/10.1145/3290605.3300268

[63] Daniel McDuff, Amy Karlson, Ashish Kapoor, Asta Roseway, and Mary Czerwinski. 2012. AffectAura: An Intelligent System for Emotional Memory. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. ACM, New York, NY, USA, 849–858. https://doi.org/10.1145/2207676.2208525

[64] Luca Menghini, Evelyn Gianfranchi, Nicola Cellini, Elisabetta Patron, Mariaelena Tagliabue, and Michela Sarlo. 2019. Stressing the accuracy: Wrist-worn wearable sensor validation over different conditions. *Psychophysiology* 56, 11 (July 2019). https://doi.org/10.1111/psyp.13441

[65] Nir Milstein and Ilanit Gordon. 2020. Validating Measures of Electrodermal Activity and Heart Rate Variability Derived From the Empatica E4 Utilized in Research Settings That Involve Interactive Dyadic States. *Frontiers in Behavioral Neuroscience* 14 (2020), 148. https://doi.org/10.3389/fnbeh.2020.00148

[66] Pejman Mirza-Babaei, Lennart E. Nacke, John Gregory, Nick Collins, and Geraldine Fitzpatrick. 2013. How Does It Play Better? Exploring User Testing and Biometric Storyboards in Games User Research. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 1499–1508. https://doi.org/10.1145/2470654.2466200

[67] Lennart Erik Nacke, Michael Kalyn, Calvin Lough, and Regan Lee Mandryk. 2011. Biofeedback Game Design: Using Direct and Indirect Physiological Control to Enhance Game Interaction. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 103–112. https://doi.org/10.1145/1978942.1978958

[68] Aditya Shekhar Nittala, Arshad Khan, Klaus Kruttwig, Tobias Kraus, and Jürgen Steimle. 2020. PhysioSkin: Rapid Fabrication of Skin-Conformal Physiological Interfaces. In *Proc. of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, New York, NY, USA, 1–10. https://doi.org/10.1145/3313831.3376366

[69] Karen Niven and Eleanor Miles. 2013. *Affect Arousal*. Springer New York, New York, NY, 50–52. https://doi.org/10.1007/978-1-4419-1005-9_1089

[70] Nami Ogawa, Takuji Narumi, Hideaki Kuzuoka, and Michitaka Hirose. 2020. Do You Feel Like Passing Through Walls?: Effect of Self-Avatar Appearance on Facilitating Realistic Behavior in Virtual Environments. In *Proc. of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376562

[71] S. Ollander, C. Godin, A. Campagne, and S. Charbonnier. 2016. A comparison of wearable and stationary sensors for stress detection. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 004362–004366. https://doi.org/10.1109/SMC.2016.7844917

[72] Fred G. W. C. Paas and Jeroen J. G. Van Merriënboer. 1994. Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review* 6, 4 (Dec. 1994), 351–371. https://doi.org/10.1007/bf02213420

[73] Matthew K.X.J. Pan, Jih-Shiang Chang, Gokhan H. Himmetoglu, AJung Moon, Thomas W. Hazelton, Karon E. MacLean, and Elizabeth A. Croft. 2011. Now Where Was I? Physiologically-Triggered Bookmarking. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 363–372. https://doi.org/10.1145/1978942.1978995

[74] Pablo E. Paredes, Francisco Ordonez, Wendy Ju, and James A. Landay. 2018. Fast & Furious: Detecting Stress with a Car Steering Wheel. In *Proc. of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174239

[75] Andrew F. H. Payne, Anne M. Schell, and Michael E. Dawson. 2016. Lapses in skin conductance responding across anatomical sites: Comparison of fingers, feet, forehead, and wrist. *Psychophysiology* 53, 7 (2016), 1084–1092. https://doi.org/10.1111/psyp.12643

[76] Frank Petzke and Daniel J. Clauw. 2000. Sympathetic nervous system function in fibromyalgia. *Current Rheumatology Reports* 2, 2 (01 Mar 2000), 116–123. https://doi.org/10.1007/s11926-000-0051-5

[77] Hugo F. Posada-Quintero and Ki H. Chon. 2020. Innovations in electrodermal activity data collection and signal processing: A systematic review. *Sensors (Switzerland)* 20, 2 (2020). https://doi.org/10.3390/s20020479

[78] Susanne Putze, Dmitry Alexandrovsky, Felix Putze, Sebastian Höffner, Jan David Smeddinck, and Rainer Malaka. 2020. Breaking The Experience: Effects of Questionnaires in VR User Studies. In *Proc. of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376144

[79] Nimesha Ranasinghe, Pravar Jain, Nguyen Thi Ngoc Tram, Koon Chuan Raymond Koh, David Tolley, Shienny Karwita, Lin Lien-Ya, Yan Liangkun, Kala Shamaiah, Chow Eason Wai Tung, Ching Chiuan Yen, and Ellen Yi-Luen Do. 2018. Season Traveller: Multisensory Narration for Enhancing the Virtual Reality Experience. In *Proc. of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174151

[80] Jean-Claude Roy, Wolfram Boucsein, Don C. Fowles, and John H. Gruzelier (Eds.). 1993. *Progress in Electrodermal Research*. Springer US. https://doi.org/10.1007/978-1-4615-2864-7

[81] Amanda Sargent, Jan Watson, Yigit Topoglu, Hongjun Ye, Rajneesh Suri, and Hasan Ayaz. 2020. Impact of Tea and Coffee Consumption on Cognitive Performance: An fNIRS and EDA Study. *Applied Sciences* 10, 7 (April 2020), 2390. https://doi.org/10.3390/app10072390

[82] Peter Schulz, Dmitry Alexandrovsky, Felix Putze, Rainer Malaka, and Johannes Schöning. 2019. The Role of Physical Props in VR Climbing Environments. In *Proc. of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300413

[83] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. 2010. Discriminating Stress From Cognitive Load

Using a Wearable EDA Device. *IEEE Transactions on Information Technology in Biomedicine* 14, 2 (March 2010), 410–417. https://doi.org/10.1109/TITB.2009.2036164

[84] David A.T. Siddle. 1991. Orienting, Habituation, and Resource Allocation: An Associative Analysis. *Psychophysiology* 28, 3 (1991), 245–259. https://doi.org/10.1111/j.1469-8986.1991.tb02190.x

[85] Erin T. Solovey, Marin Zec, Enrique Abdon Garcia Perez, Bryan Reimer, and Bruce Mehler. 2014. Classifying Driver Workload Using Physiological and Driving Performance Data: Two Field Studies. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 4057–4066. https://doi.org/10.1145/2556288.2557068

[86] Raphaela Stadler, Allan Stewart Jepson, and Emma Harriet Wood. 2018. Electrodermal activity measurement within a qualitative methodology. *International Journal of Contemporary Hospitality Management* 30, 11 (Nov. 2018), 3363–3385. https://doi.org/10.1108/ijchm-11-2017-0781

[87] Chiew Seng Sean Tan, Johannes Schöning, Kris Luyten, and Karin Coninx. 2014. Investigating the Effects of Using Biofeedback as Visual Stress Indicator during Video-Mediated Collaboration. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 71–80. https://doi.org/10.1145/2556288.2557038

[88] Yigit Topoglu, Jan Watson, Rajneesh Suri, and Hasan Ayaz. 2019. Electrodermal Activity in Ambulatory Settings: A Narrative Review of Literature. In *Advances in Neuroergonomics and Cognitive Engineering*. Springer International Publishing, 91–102. https://doi.org/10.1007/978-3-030-20473-0_10

[89] Panagiotis Tsiamyrtzis, Malcolm Dcosta, Dvijesh Shastri, Eswar Prasad, and Ioannis Pavlidis. 2016. Delineating the Operational Envelope of Mobile and Conventional EDA Sensing on Key Body Locations. In *Proc. of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 5665–5674. https://doi.org/10.1145/2858036.2858536

[90] Graham Turpin, Phil Shine, and Malcolm Lader. 1983. Ambulatory Electrodermal Monitoring: Effects of Ambient Temperature, General Activity, Electrolyte Media, and Length of Recording. *Psychophysiology* 20, 2 (March 1983), 219–224. https://doi.org/10.1111/j.1469-8986.1983.tb03291.x

[91] Róisín Vahey and Rodrigo Becerra. 2015. Galvanic skin response in mood disorders: A critical review. *International Journal of Psychology & Psychological Therapy* 15, 2 (2015), 275–304.

[92] Marieke van Dooren, J.J.G. (Gert-Jan) de Vries, and Joris H. Janssen. 2012. Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & Behavior* 106, 2 (May 2012), 298–304. https://doi.org/10.1016/j.physbeh.2012.01.020

[93] Hendrika G. van Lier, Marcel E. Pieterse, Ainara Garde, Marloes G. Postel, Hein A. de Haan, Miriam M. R. Vollenbroek-Hutten, Jan Maarten Schraagen, and Matthijs L. Noordzij. 2019. A standardized validity assessment protocol for physiological signals from wearable technology: Methodological underpinnings and an application to the E4 biosensor. *Behavior Research Methods* 52, 2 (July 2019), 607–629. https://doi.org/10.3758/s13428-019-01263-9

[94] Pieter Vanneste, Annelies Raes, Jessica Morton, Klaas Bombeke, Bram B. Van Acker, Charlotte Larmuseau, Fien Depaepe, and Wim Van den Noortgate. 2020. Towards measuring cognitive load through multimodal physiological data. *Cognition, Technology & Work* (12 Jul 2020). https://doi.org/10.1007/s10111-020-00641-0

[95] Vygandas Šimbelis, Anders Lundström, Kristina Höök, Jordi Solsona, and Vincent Lewandowski. 2014. Metaphone: Machine Aesthetics Meets Interaction Design. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 1–10. https://doi.org/10.1145/2556288.2557152

[96] Chat Wacharamanotham, Lukas Eisenring, Steve Haroz, and Florian Echtler. 2020. Transparency of CHI Research Artifacts: Results of a Self-Reported Survey. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376448

[97] Chat Wacharamanotham, Matthew Kay, Steve Haroz, Shion Guha, and Pierre Dragicevic. 2018. Special Interest Group on Transparent Statistics Guidelines. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/3170427.3185374

[98] Günter Wallner, Nour Halabi, and Pejman Mirza-Babaei. 2019. Aggregated Visualization of Playtesting Data. In *Proc. of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300593

[99] Chen Wang, Erik N. Geelhoed, Phil P. Stenton, and Pablo Cesar. 2014. Sensing a Live Audience. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 1909–1912. https://doi.org/10.1145/2556288.2557154