Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies

Ana María Rojo López
University of Murcia, Spain
anarojo@um.es
https://orcid.org/0000-0003-4303-9047

Paweł Korpal
Adam Mickiewicz University, Poznań (Poland)
pkorpal@amu.edu.pl
https://orcid.org/0000-0001-7479-936X

Abstract

This article offers a critical appraisal of two experimental methods used to provide physiological measures of stress and emotions in translation and interpreting research, namely, the analysis of heart rate and heart rate variability, and skin conductance. This is a hands-on introduction to summarize information for fellow-researchers on what these methods are and what they tell us about our body and mind as well as to offer a comprehensive summary of practical applications and analysis standards. The first part of the article introduces the ways in which emotions are experienced and processed in the brain; it provides a framework for interpreting physiological arousal and its role in the perception and construction of emotions. The second part is structured in two parallel sections devoted to each of the two experimental methods. Both sections review existing research on these methods in Translation and Interpreting Studies and discuss the way in which each can best be used in experimental research. They offer suggestions on experiment planning, measurement, data analysis and data reporting. A final remark on ethics and triangulation is offered and some emerging challenges are addressed.

Keywords: heart rate; heart rate variability; skin conductance; emotion; stress

1. Introduction

The focus of Cognitive Translation and Interpreting Studies (CTIS) on the translator's mind has benefitted from scholars' engagement with the theoretical and methodological developments of other disciplines, such as psychology and other areas related to cognitive science (Halverson,

2010). From early computational approaches to more current embodied, embedded, enacted, extended, affective (4EA) cognition approaches (Muñoz Martín, 2010), the evolution of CTIS has been marked by the scholars' substantiated efforts towards theory-building and methodological innovation. But what started as a timid attempt has resulted in a technological boom that is nowadays viewed by many as a threat to the reliability and ecological validity of research. The increased sophistication of experimental designs has moved participants from computer screens and keyboards to experimental settings bearing no resemblance whatsoever to a likely translation workplace. Professional and student translators have consented to work while having their eye movements read by an eye-tracking device (Hvelplund, 2014) or wearing electroencephalography (EEG) caps with sensors wired up to a computer (Grabner et al., 2007). Interpreters have:

- had samples of their saliva taken in order to measure levels of cortisol (e.g., AIIC, 2002);
- been asked to wear pulsometer watches with bands under their chests (e.g., Korpal, 2016; Rojo López et al., 2014) or galvanic skin sensors (e.g., Korpal & Jasielska, 2019); and
- even been asked to interpret or translate while lying inside a functional magnetic resonance imaging (fMRI) scanner (e.g., Hervais-Adelman et al., 2015; Szpak et al., in press).

Such an unprecedented surge of interest in technology has certainly advanced our knowledge of what is going on in the translator's mind and body while they are performing a task. However, concerns have also been raised about replacing the more human aspects of translating with the uncontrolled use of technological gadgets. The debate continues and both sides could have a case. On the one hand, there is an urgent need to define measurement and analysis parameters for applying these instruments to translation and interpreting environments. On the other, research has already started to bear fruit and can now provide elements to inform the debate. The time is now ripe for a critical evaluation of experimental methods and technological advantages.

This article aims to contribute to the existing debate by providing a critical appraisal of two experimental methods used to provide physiological measures of stress and emotions in translation and interpreting (T&I) research, namely, the analysis of heart rate (HR) and heart rate variability (HRV), and of skin conductance (SC). The purpose of the article is to provide a hands-on introduction that tells researchers what these methods are and what they tell us about our body and mind; it also aims to provide a comprehensive summary of practical applications and analysis standards. Existing research is examined and tips for future research are suggested. Before reviewing the research based on these methods and discussing the ways they can be used in experimental research, we should first sketch the manner in which emotions are experienced and processed in the brain.

2. How do we experience emotions and how are they processed in the brain?

In defining how emotions are experienced and processed in the brain, two competing models can be distinguished: the classical view of emotion and the theory of constructed emotion (Barrett, 2017b). Even if these views coincide with the physiological signs of emotional experience or arousal, they hold fundamentally different perspectives on the ways in which emotions are processed and constructed in the brain.

The classic model envisages emotions as response patterns consisting of three types of components – overt behaviours, autonomic responses and hormonal secretions – controlled by separate neural systems (Ekman, 1992; Frijda & Scherer, 2009). This model rests on the assumption that the human body is operated by two distinct nervous systems – the central and the peripheral – that work in close interaction. The central nervous system (CNS) consists of the brain and the spinal cord, whereas the peripheral nervous system consists of all the nerve tissue that branches from them. Everyday experiences are full of examples of the interaction between both systems: when, for instance, we start to develop beads of sweat (a physiological reaction controlled by the peripheral system) when we cannot remember the answers to a test (a cognitive function controlled by the central system). The peripheral nervous system can be also divided into the autonomic nervous system (ANS) – which controls inner organs, such as our heart – and the somatic or voluntary nervous system, which is in charge of body and muscle movements. In turn, the ANS is made up of the sympathetic (SNS) and the parasympathetic nervous systems (PNS). These two systems also interplay and regulate each other, acting somehow as an acceleration and a braking system: while the SNS increases physiological arousal when active, the PNS slows it down when active. So, when we are stressed, the SNS is activated and the PNS is inhibited; when we are calm, the PNS is more activated while the SNS remains less so.

According to the classical view, emotional response starts in the limbic system, the part of the brain specifically equipped for emotional processing (Davis, 1992; LeDoux, 1991). For instance, when we experience a potentially dangerous or harmful situation, our eyes and ears send the information to the amygdala, the part of the limbic system in charge of emotional evaluation. When danger is perceived, a distress signal is then sent to the hypothalamus, the part in charge of regulating our response to emotions. The hypothalamus first activates the SNS by sending signals to the adrenal glands through the nerves. These glands then pump the hormone *adrenaline* into the bloodstream, causing a number of physiological changes, such as the acceleration of heart beat and breathing. Extra oxygen is sent to the brain and all our senses become sharper, causing alertness and preparing the fight-or-flight response (Cannon & Rosenberg, 1932). These symptoms are so obvious that we not only believe we can identify them when they happen to us, but we can also identify them in others fairly accurately.

As intuitive and down to earth as this view may seem, this model faces three fundamental contradictions that are at the roots of the *emotion paradox* – that is, they seem self-evident and real and yet they defy clear, scientific definition (Barrett, 2006). First, the model assumes that

different categories of emotion – anger, sadness, fear, disgust, happiness, surprise – have their own physiological fingerprints; that is, that they create specific combinations of facial expression, body language and other physiological cues that allow us to recognize instances of them in others, and experience them ourselves. Second, it presupposes that emotions are biologically wired in our brains from birth, so we have circuits devoted to specific emotions and that they cause these specific sets of changes or fingerprints – what Lindquist and her colleagues term "locationism" (see Lindquist et al., 2012). Third, it still rests on the assumption that the mind and the body are separate entities (the CNS and the ANS) linked by a cause-and-effect relationship. This has given rise to an unresolved debate between "peripheral" theories of emotion which posit that physiological reactions precede emotions and "central" theories which point to specific brain circuits that determine emotions (Pace-Schott et al., 2019, p. 268).

Empirical evidence from psychophysiology and neurology has failed to provide support for the existence of emotions as discrete categories of experience with a unique fingerprint processed or originated in specific regions of the brain. Recent neuroscientific evidence suggests that what exists in the brain and the body is affect. Emotions, rather than being genetically programmed, are something that our brain constructs on the basis of the dynamic interaction of various components, in which "emotions generate efferent influences on peripheral physiology from central representations (including predictive active inference) and are, in turn, shaped by afferent information from the periphery (including prediction errors)" (Pace-Schott et al., 2019, p. 268). To solve this paradox, the theory of constructed emotion rebuffs the idea that there are specific emotional circuits in the brain; instead, it suggests that emotions are constructed by using multiple brain networks working together. It departs from the classical model's distinction between the CNS and the peripheral nervous system to propose a more holistic model where emotions are envisaged as "whole brain–body phenomena in context" (Barrett, 2017a, p. 16). In this new model, the barrier between cognitive functions and physiological reactions is broken down by proposing a new brain system guided by two general principles – allostasis and interoception – that work in unison, looping through various brain regions (Kleckner et al., 2017). *Interoception* represents sensations from within the body whereas *allostasis* (or *body budgeting*) regulates the peripheral systems according to costs and benefits. Together, both processes keep our body systems balanced as we respond to internal and external stressors. Achieving this balance with efficiency requires the ability to anticipate the body's needs and even satisfy them before they arise. To do this, the brain runs a mental model of the body in the world (what in psychology is referred to as *embodied simulation*), which includes the relevant statistical regularities both of the outside world and of the internal milieu (i.e., interoception). Therefore, when it receives a new sensory input, the brain starts a process to interpret (or categorize) it by comparing it with the available sensory array, and it starts to guide action in relation to energy costs and potential rewards for the body.

In this model, emotions are constructed or interpreted as the rest of perceptions by using the same neuro-anatomical principles that govern information flow in the brain – in biology, the principle predicting that the same outcome may be produced in multiple ways is called

degeneracy (Barrett, 2017b). From this point of view, all the symptoms of emotional arousal have a physical function: for example, the main function of a fast-beating heart is to pump enough oxygen into our limbs so that we can run. But how do we give this sensation additional meaning and categorize it as an emotional experience? When we experience affect with high arousal and unpleasant valence of the kind previously described, the meaning we make from it depends on how we categorize it: it may be an emotional instance of fear, but it may also be a physical reaction to a high intake of caffeine or even to the perception that your friend is a bully. Our *interoceptive network* will create multiple competing interpretations but, in order to determine which prediction is the winner, it will need the assistance of the *control network.* This network "controls" the process by working as a kind of an optimizer which directs attention towards relevant sensory input or more intense affect and makes some predictions fit while others are discarded as irrelevant (Barrett, 2017b, p. 123). As a result, instances most suitable to the current situation or environment will be selected to shape our perception and action (e.g., fear will be the most salient interpretation of the experienced affect when a person watches a horror film, and we may accordingly choose to close our eyes to stop perceiving the distressing visual input). In brief, our brain constructs emotions in the same way that it constructs meaning, that is,

> by anticipating (predicting and adjusting to) incoming sensations. Sensations are categorized so that they are (i) actionable in a situated way and therefore (ii) meaningful, based on past experience. When past experiences of emotion (e.g. happiness) are used to categorize the predicted sensory array and guide action, then one experiences or perceives that emotion (happiness). (Barrett, 2017a, p. 9)

The theory of constructed emotion has significant implications for experimental research, some of which are discussed in this article. If patterns of physiological arousal cannot be unequivocally matched with specific emotions, then emotions cannot be measured on the basis of physiological responses, at least not exclusively so. Data from physiological changes should always be triangulated with behavioural data and with data from self-reports on the subjects' perception of emotions in specific contexts. As Barrett (2017b) outlines "instead of building one model that reads anger for everybody in the whole world, you may need to build a model that reads the variable changes in anger for a single person over time, depending on the context" (p. 204). We agree with constructivist models that emotion and stress should no longer be considered independent phenomena but rather two sides of the same "perceptual" coin. Constructed via the same brain mechanisms, the only difference between emotion and stress is found in the end-result, in "whether your brain categorizes your sensations as stressful or emotional" (Barrett, 2017b, p. 204).

We have provided a more holistic and dynamic framework for interpreting physiological arousal and its role in the perception and construction of emotions. In what follows, we aim to provide a common and more robust framework for measuring and analysing two indices interpreted in experimental T&I research as reflecting translators' and interpreters'

Rojo López, A. M., & Korpal, P. (2020). Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 191–217.

physiological states and emotional arousal: HR and HRV, and SC. We discuss the former first, in Section 3.

3. Heart rate and heart rate variability

Both HR and HRV have received much attention recently as indices of psychological and physical well-being. We would like to outline the scientific rationale behind each one, and also their advantages and disadvantages, so as to foster correct use of these metrics.

HR focuses on the average number of beats in a given period of time, but it does not imply a regular rhythm between heartbeats. Therefore, an HR of 60 beats per minute could mean a regular rhythm of 1 beat per second or it could mean a 60-beat count whose interspersed lapses lasted, say, 0.5 s, 1.5 s, 0.5 s, 1.5 s. Its main advantages are that it is easy to measure by a variety of non-invasive devices and wearables, and does not require extreme accuracy. One of its greatest disadvantages – at least for experiments that do not involve physical exercise – is that it focuses mainly on cardiovascular activity and, therefore, works best to measure exertion during exercise, that is, as a measure of physical fitness; while at rest, it is only a vague indicator of internal activity or emotional regulation. Because of its poor accuracy, most studies that focus on emotional regulation use HR in combination with other methods, or have started to use HRV instead.

HRV measures the specific changes (or variability) between successive heart beats, accounting for the contribution of the nervous system – more specifically of the *vagus* nerve, hence the label *cardiac vagal tone* – to cardiac regulation. HRV is, therefore, a more complex and holistic measurement of *autonomic nervous system* activity, which integrates the nervous, cardiovascular and respiratory systems. HRV is thus preferred over HR when the focus is on vagal tone and its correlation with better executive cognitive performance, as well as better emotional and health regulation (cf. Laborde et al., 2017). At the risk of oversimplification, a decrease in HRV – that is, less variability between heartbeats – can often be assumed to indicate that the body is under stress from exercise, psychological events, or other internal or external stressors. Conversely, an increase of HRV usually points to a greater ability to tolerate stress or it suggests recovery from prior accumulated stress (Kim et al., 2018). In contrast to HR, HRV needs greater accuracy and requires the use of electrocardiograph (ECG) technology or more sophisticated wearable monitors. It is also best measured during a resting state – a problem for using wrist monitors in Translation Studies due to the repetitive actions involved in keyboarding – with high HRV being considered generally favourable and low HRV as unfavourable. However, the opposite holds true when in an active state, in which lower relative HRV is generally more favourable than higher HRV.

3.1. HR and HRV in translation and interpreting

In the past two decades, the use of HRV as a measure of cardiac vagal tone has grown exponentially in psychophysiological research to measure self-regulation at the cognitive,

emotional, social and health levels (Laborde et al., 2017). This increasing popularity has been mainly due to some of the reasons mentioned above: it is a non-invasive and pain-free method; it is rather economical, compared to more sophisticated methods; and recent progress in technology has made data collection and analysis easily accessible (e.g., with bluetooth HR monitors and mobile applications).

Translation and interpreting researchers have also responded to these compelling advantages, timidly starting to incorporate HRV into studies that focus on psycho-affective factors. To our knowledge, the majority of existing T&I studies use HR rather than HRV as a physiological manifestation of stress in combination with other physiological measures, such as blood pressure (BP) alone (Klonowicz, 1994; Korpal, 2016), BP and SC level (Kurz, 2002, 2003) or salivary cortisol (AIIC, 2002) or even in conjunction with behavioural measures such as the log reaction time of key press (Rojo López & Ramos Caro, 2014). Most of these studies are found in interpreting, audiovisual translation (AVT) and media accessibility, the last of these being mainly reception studies designed to test the audience's emotional response and level of immersion in film viewing. The only work based on written translation is also a reception study exploring the role of readers' emotional responses to metaphorical versus non-metaphorical target-language (TL) translations of source-language (SL) metaphors of emotion (Rojo López et al., 2014). This lack of HR and HRV studies in written translation is probably due to task and equipment requirements – that is, typing requires the continuous movement of wrists and hands whereas the measurement of HRV demands being as still as possible.

In Interpreting Studies, some of the factors that have an effect on the interpreters' level of stress by using HR are the stressful nature of conference and simultaneous interpreting (AIIC, 2002; Klonowicz, 1994; Kurz, 2003); different tasks or situations – such as media vs on-site interpreting (Kurz, 2002; Roziner & Shlesinger, 2010) or being on-mike as opposed to being off-mike, helping the active interpreter (AIIC, 2002); and the speed of the speaker's delivery (Korpal, 2016, 2017). To our knowledge, only one recent study on interpreting uses HRV to explore the interplay between stress and cognitive control while performing two different tasks: shadowing in L1 (Russian) and L2 (German/English) and simultaneous interpreting to and from the L2 (Chernigovskaya et al., 2019).

Studies on AVT and media accessibility have used HR mainly to test the audience's emotional response to AV material, whether to different audio-described versions of emotionally loaded video clips, such as a more objective versus a more subjective audio description (Ramos Caro, 2015; 2016), or to different AVT modalities, as in dubbing versus voice-over (Iturregui-Gallardo et al., 2018).

A quick look at the methodology of these studies for HR analysis suffices to realize that there is a need for greater rigour and consistency. More often than not, analysis and measurement methods are not explained in enough detail and there seems to be no agreement between experimental designs, measurement times and even the variables selected for HR analysis.

This kind of experimental lassitude is becoming a growing concern among T&I researchers, who have recently started to raise their voices to ask for the definition of common frameworks favouring better practice for experimental research. Some researchers justly reproach T&I researchers for not implementing experimental methods with the same rigour as in the disciplines from which such methods were borrowed (Orero et al., 2018).

Even in psychophysiology, a pressing need has been declared for scholars to become aware of key methodological issues that should be considered for analysing HRV. Ever since the initial recommendations of the Task Force on HRV for measurement and interpretation issues (Malik, 1996), methodological issues have been discussed in a series of works culminating with the recommendations by Laborde et al. (2017) to conduct a full research project with HRV. These recommendations have served as the basis for the suggestions offered in the present section on experiment planning, measurement, data analysis and data reporting. Given the current theoretical focus on HRV and variables reflecting vagal tone, the next section shines the spotlight on them.

3.2. Which HRV variables should be analysed?

Obviously, the answer to this question will depend on the researcher's interest and research questions. More than 70 variables can be calculated from HRV analysis, so some information about the variables it reflects can be of use when trying to select those that are appropriate.

HRV measures can be classified into two types: one based on HR (*time domain measures*) and the other based on interbeat interval (*frequency domain measures*). Time domain measures (see Figure 1) record time variations between consecutive R waves on an ECG. An R wave in the normal surface electrocardiogram is the initial upward deflection of the QRS complex, following the Q wave, and it represents ventricular depolarization (activation). In other words, it is the main spike seen on an ECG line.

Time domain measures are most frequently based on three parameters:

- the standard deviation of normal intervals between successive Rs (i.e., RRIs) (SDNN), which reflects all the cyclic components responsible for variability in the period of recording;
- the root mean square of normal to normal interval differences (RMSSD), which reflects vagal tone and correlates highly with high-frequency parameters – although it differs from these in that it is relatively free of respiratory influences;
- the proportion of RRIs of more than 50 ms (pNN50), which also reflects vagal tone and correlates with RMSSD and HF.

In addition, vagal tone can be inferred by the peak–valley analysis, which acts as a time–domain filter dynamically focused at the exact ongoing respiratory frequency.
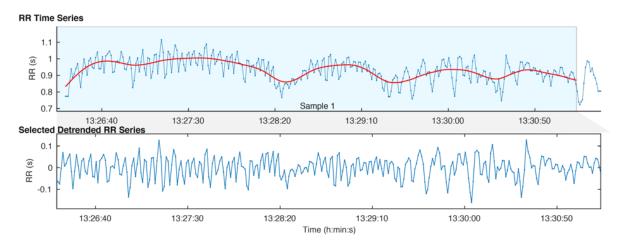

Figure 1. An example of RR time series measures from *Kubios HRV Premium*

Frequency domain measures take into account "the decomposition of the wave form of R-R or interbeat intervals (RRIs or IBI) into frequency power bands using spectral analysis" (Bassett, 2016, p. 512). The following frequency bands can be distinguished:

- *ultra-low* frequency (ULF <0.0033 Hz), which requires 24 h recordings and reflects circadian oscillations, core body temperature, metabolism and the renin-angiotensin system – that is, the system regulating BP, fluid balance and cardiovascular resistance;
- *very-low* frequency (VLF = 0.0033–0.04 Hz), which represents long-term regulation mechanisms, thermoregulation and hormonal mechanisms (Malik, 1996);
- *low*-frequency (LF = 0.04–0.15 Hz), which represents a mix between sympathetic and vagal influences;
- *high*-frequency (HF = 0.15–0.4 Hz), which reflects vagal tone when breathing rates remain between nine cycles per minute (0.15 Hz) and up to 24 cycles per minute (0.40 Hz), and HRV stays between these frequencies. The HF band is actually frequently called the "respiratory band" because it corresponds to the HR variations related to the respiratory cycle.

Researchers should take into consideration that frequency bands may need to be adjusted to the sample population – for example, children and infants breathe faster – and their range should be moved to 0.24–1.04 Hz at rest. Two important recommendations for practice are, therefore, to take previous research measures as reference or to calculate the respiratory rates of the sample, and to couple frequency analysis with other time–domain parameters that indicate vagal tone and are less affected by breathing, as is the case with RMSSD.

Finally, the LF to HF ratio has also long been considered a frequency-domain HRV measure representing the balance between the sympathetic and parasympathetic systems. However,

Rojo López, A. M., & Korpal, P. (2020). Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 191–217.

this view is now highly controversial, owing to the strong evidence against the relationship between LF power and sympathetic nerve activation and also against the linear and reciprocal relationship between sympathetic and parasympathetic nerve activity (Heathers, 2014). For this reason, Laborde et al. (2017, p. 5) recommend that researchers use indices that clearly reflect physiological systems with a theoretical underpinning, such as the indices of vagal tone (i.e., RMSSD, peak–valley and HF). See Figure 2 for results of time–domain vs frequency–domain measures.
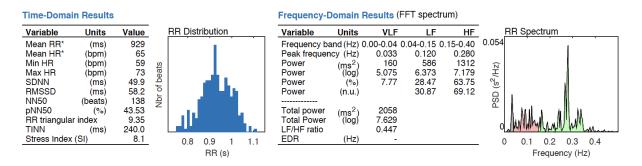
Time-Domain Results

Variable	Units	Value
Mean RR*	(ms)	929
Mean HR*	(bpm)	65
Min HR	(bpm)	59
Max HR	(bpm)	73
SDNN	(ms)	49.9
RMSSD	(ms)	58.2
NN50	(beats)	138
pNN50	(%)	43.53
RR triangular index		9.35
TINN	(ms)	240.0
Stress Index (SI)		8.1

Frequency-Domain Results (FFT spectrum)

Variable	Units	VLF	LF	HF
Frequency band (Hz)		0.00-0.04	0.04-0.15	0.15-0.40
Peak frequency	(Hz)	0.033	0.120	0.280
Power	(ms^2)	160	586	1312
Power	(log)	5.075	6.373	7.179
Power	(%)	7.77	28.47	63.75
Power	(n.u.)		30.87	69.12

Total power	(ms^2)	2058		
Total Power	(log)	7.629		
LF/HF ratio		0.447		
EDR	(Hz)	-		

Figure 2. An example of time–domain and frequency–domain results from *Kubios HRV Premium*

Apart from these measures, there are also some non-linear indices based on the interbeat interval (see Figure 3). One of these non-linear indices is the Poincaré plot, a type of recurrence plot used to measure self-similarity in different processes. In the context of HRV, the plot illustrates quantitative and qualitative patterns of one's individual HRV in the shape of an ellipse, by plotting each R–R interval as a function of the previous one. Two different standard deviations – which result from the orthogonal distances between the scatter and the elliptical diameters – are also added to the ellipse: crosswise (SD1) and lengthwise (SD2). Whereas SD1 is considered more sensitive to quick and high-frequent changes, SD2 is regarded as an index of long-term changes. Piskorski and Guzik (2005) consider this type of non-linear measure to be more adequate than linear indices to measure the complex and erratic fluctuations of the ANS. Poincaré plots are also claimed to be indicators of vagal activity and reduced cardiac vagal control associated not only with physiological, but also with psychological strain and stress (Melillo et al., 2011). However, Laborde et al. (2017) advise using them as complementary indices of HRV because their potential to predict psychophysiological phenomena still requires further demonstration.
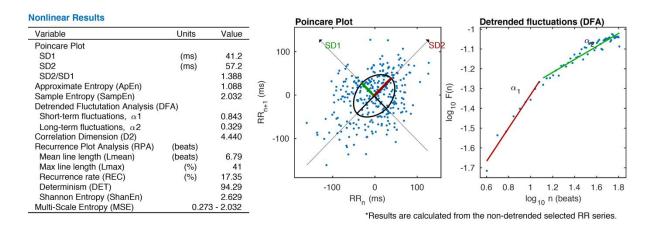
Rojo López, A. M., & Korpal, P. (2020). Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 191–217.

Nonlinear Results

Variable	Units	Value
Poincare Plot		
SD1	(ms)	41.2
SD2	(ms)	57.2
SD2/SD1		1.388
Approximate Entropy (ApEn)		1.088
Sample Entropy (SampEn)		2.032
Detrended Fluctutation Analysis (DFA)		
Short-term fluctuations, $\alpha 1$		0.843
Long-term fluctuations, $\alpha 2$		0.329
Correlation Dimension (D2)		4.440
Recurrence Plot Analysis (RPA)	(beats)	
Mean line length (Lmean)	(beats)	6.79
Max line length (Lmax)	(%)	41
Recurrence rate (REC)	(%)	17.35
Determinism (DET)		94.29
Shannon Entropy (ShanEn)		2.629
Multi-Scale Entropy (MSE)		0.273 - 2.032

*Results are calculated from the non-detrended selected RR series.

Figure 3. An example of non-linear results from *Kubios HRV Premium*

3.3. Should HRV be treated as a dependent or an independent variable?

In T&I studies, HRV is most likely to be treated as a dependent variable analysed to explore the way in which it differentiates groups assigned to different experimental conditions (e.g., performing different tasks or receiving different versions of the same stimuli). It may also be frequently related to other subjective indices of stress or emotions – for example, those provided by the State-Trait Anxiety Inventory (STAI; Spielberger et al., 1983), the Self-Assessment Manikin (SAM) questionnaire (Bradley & Lang, 1994) or even tailor-made questionnaires on emotional perception – by using correlations and regression analyses.

However, there is also the possibility of treating HRV as an independent variable to explore, for instance, the relationship between resting HRV as an individual difference and task performance. As Laborde et al. (2017) outline, justification for this possibility can be found on existing results that consistently associate high resting HRV with positive outcomes. Participants could therefore be assigned to two different groups divided by median split into high and low RMSSD.

3.4. Which confounding variables should be controlled?

Once the experimental variables have been selected, researchers should also decide on which confounding variables influencing HRV need to be controlled. Several factors that affect ANS activity may act as confounding variables: age and gender, physical activity, alcohol and nicotine consumption, food and water intake, circadian rhythm, HR, BP, respiration and even several medications (Bassett, 2016). Laborde et al. (2017) list a series of stable and transient variables recommended for controlling in HRV studies. Among stable variables, they include:

- age and gender;
- smoking;
- habitual alcohol consumption;

- waist, height and waist-to-hip ratio;
- cardioactive medication (i.e., antidepressant, antipsychotic and antihypertensive); and
- oral contraceptive intake for female participants.

The last of these may not influence HRV at rest, but it may affect response to stressful conditions.

Laborde et al. (2017) underline the importance of recording confounding variables for excluding participants and interpreting outliers or anomalous data points but, whenever possible, they recommend obtaining objective measures of these potentially confounding factors. In order to control for transient variables, the following recommendations are offered:

- the day before the experiment, the participants should follow a normal sleep routine (typical bed time should be recorded) and avoid both intense physical training and alcohol consumption;
- in the two hours before the experiment, no meal, coffee, tea or caffeinated or energizing drinks should be consumed;
- finally, just before the experiment begins, the participants should be asked if they need to use the bathroom.

3.5. How many participants and comparison groups are needed?

One of the recurrent quandaries researchers face when planning a research project is deciding how many participants and groups are needed for statistical power and representation of the target population.

Regarding power, effect and sample size, Laborde et al. (2017) use Quintana's (2016) distribution analysis of 300 HRV effect sizes to warn that HRV studies are frequently underpowered; and they recommend to adapt the interpretation of Cohen's effect sizes from 0.20, 0.50 and 0.80 to 0.25, 0.50 and 0.90 as, respectively, small, medium and large effect sizes. Similarly, Quintana's analysis also suggests that to achieve 80 per cent power, samples of 233, 61 and 21 participants are needed to detect, respectively, small, medium and large effect sizes. Nevertheless, whenever different effect sizes or statistical powers are needed, researchers are recommended to use *a priori* procedures to compromise on the sample size and carefully calculate their margin of error and confidence interval, or even to resort to statistical packages (e.g., G*Power 3 or pwr) that allow researchers to calculate the sample size required.

Sample size is closely related to selecting the best experimental design for HRV measurement: Should we opt for a within-subject or a between-subject comparison? Owing to the variable and complex nature of HRV, Laborde et al. (2017) point to the following advantages of within-subject over between-subject designs: they allow for greater experimental control; individual

Rojo López, A. M., & Korpal, P. (2020). Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 191–217.

differences in respiratory rates are easier to monitor and control; given their increased statistical power, fewer participants are required and the impact of external factors (e.g., medication, alcohol, smoking) is minimized. Nonetheless, when the experiment needs to be carried out on different days, they recommended that researchers keep the time of day constant and use non-identical correlated tasks that investigate similar constructs to avoid habituation to experimental conditions.

3.6. Which experiment structure is recommended?

T&I experiments show a generalized lack of agreement regarding the number of HR measures taken. While some take baseline measures before and after an experiment (e.g., Ramos Caro, 2016), others use only one post-test measure (e.g., Korpal, 2016 recorded pre- and post-HR measures, but used the post-test measure only as a baseline value) or compute the mean HR during the whole experiment as the baseline against which HR during experimental time is contrasted later on (Rojo López et al., 2014).

Laborde et al. (2017) recommend a structure based on three measures collected at three different points in time, that is, baseline, event and post-event, which reflect what they call the three Rs of HRV: *resting, reactivity, recovery.* This structure allows for *tonic* and *phasic* HRV to be investigated. Tonic HRV is taken at one point in time; it is also called *resting* HRV. Phasic HRV represents change in HRV from two different points in time; it is also called *reactivity* or *change delta* HRV. In the proposed three R structure, tonic HRV can be measured at each of the three points (i.e., baseline, event, post-event) and phasic HRV can be measured from the change between any two points, baseline and event (i.e., "reactivity"), task and post-event (i.e., "recovery") and baseline and post-event.

3.7. What are the standards for HRV measurement?

HRV can be measured through different recording techniques: through ECG recordings using electrodes, which leads to very precise correction (see Section 3.8 for artefact correction); measuring the IBI through chest or wrist belts coupled to HR monitors; or photoplethysmography, a technique that involves shining a small light onto an area where capillaries are easy to access. Nowadays, different models of sports watch or smartphone equipped with sensors allow for HRV to be measured on the wrist. In the past decade, there has been a growing demand for these sports watches equipped with HR monitors and advanced software for analysing HRV. Even though not all the models are considered reliable, the series of Polar® RS800G3™ and Polar® S810i™ have been tested, demonstrating that they are able to capture series of RR intervals for analysing HRV indexes as reliable as those obtained by ECG (Penachini da Costa de Rezende Barbosa et al., 2014).

A relevant standard for the accuracy of HRV measurement is the sampling rate of the data-acquisition system. The recommendations of the Task Force (Malik, 1996) advised setting the sampling rate at a minimum of 200 Hz, but Laborde et al. (2017) suggest a sampling rate of at

least 125 Hz for researchers in psychophysiological research, with a more conservative lower boundary of 500 Hz.

Researchers should bear in mind that recordings of different lengths should not be compared. Standards for measurement duration follow the recommendations of the Task Force and use short-term recordings of five minutes whenever possible to enable comparison between studies. Even if – depending on the research question and the needs of an experimental design – shorter recordings can be considered, they should last at least one minute when vagal tone is targeted so as to make frequency analysis possible. Ultra-short analysis of even ten seconds for vagal tone is also possible with time-domain analysis, but justification from researchers would be in order. For long-term recordings, 24-hour indexes require controlling for physical activity and respiration. Regarding the control of respiration, the most common recommendations include monitoring it whenever possible and checking whether respiratory frequency remains between nine and 24 cycles per minute (corresponding to the HF band, 0.15–0.40 Hz). Despite the possibility of long-term recordings in ambulatory settings, though, it is important to bear in mind that for unambiguous interpretation of psychophysiological phenomena, in particular vagal tone, HRV has to be measured without any physical activity. Under any other circumstances, measurement is possible but prevents a clear interpretation of vagal tone.

Procedural standards during baseline recording are also important to guarantee comparability across studies. The body position during the experiment should be as close as possible to the one used during the baseline. The recommendations by Laborde et al. (2017, p. 9) include taking a baseline recording while sitting with your knees at an angle of 90º, both feet flat on the floor, hands on thighs, palms facing upward (to avoid interoceptive effects if participants feel their wrist pulse) and eyes closed.

3.8. What are the recommendations for HRV data analysis and reporting?

Laborde et al. (2017) provide specific recommendations for data analysis and reporting, including suggestions on which software to use and the procedure for ensuring artefact correction and normality of data. The most useful points are summarized below. The most popular HRV software is Kubios; the premium version allows users of Polar HR monitors to link to their Polar Flow account and export their training data. It also provides automatic artefact correction – that is, it allows researchers to filter their data automatically, detecting the RR intervals differing "abnormally" from the normal mean RR interval that may represent an artefact. This procedure is adequate when data are recorded in good conditions with little or no movement, but Laborde et al. (2017) warn that errors are still possible and the artefacts deleted may correspond to real heartbeats. For this reason, they recommend that the ECG signal be visually inspected to minimize the critical consequences of deleting a real heartbeat assumed to be an artefact.

Rojo López, A. M., & Korpal, P. (2020). Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 191–217.

One crucial issue in data analysis and reporting is inconsistencies regarding the variables used to analyse vagal tone, a fact that hinders the comparability of results across experiments. As reported (see Section 3.2), the three main variables recommended for exploring vagal tone are either RMSSD, peak–valley and HF. Even if only one of these variables is selected, researchers are recommended to replicate analyses with the other variables to ensure the consistency of the results. Similarly, even if studies focus on one main variable, they are also encouraged to submit all the raw data and the analysis of the other HRV parameters as supplementary material.

4. Skin conductance

SC, also referred to as galvanic skin response (GSR) or electrodermal activity (EDA), has long been used as an autonomic measure of emotion in empirical research.

Similarly to HR and HRV, SC is based on the fight-or-flight response, as sweating helps to cool down the body and, in turn, prepare the organism to cope with a stressful situation or a perceived threat (PsychLab, 2003). It is common knowledge that sweating can indeed be associated with emotional arousal. Most people have experienced this at least once, for example, when waiting for examination results or being on stage and giving a talk to 200 people. Researchers have been using this mechanism in the laboratory, assuming that electrodermal activity can inform us about both physiological and psychological arousal. The main assumption behind measuring electrodermal activity in the laboratory is that higher SC values index an increased intensity of emotion experienced by a study participant (Bradley et al., 1990; Cook et al., 1991; Waugh et al., 2011).

In order to record SC in the laboratory, an electrodermal activity amplifier is necessary. This might be accompanied by a modular data-acquisition unit that makes it possible to record signals simultaneously from other channels, such as ECG or electromyography (EMG). To measure exosomatic EDA, either disposable (single-use) or reusable electrodes can be used. Single-use electrodes might be pre-gelled, while in the case of reusable electrodes an isotonic gel is applied in the electrode's gel cavity. A small voltage of 0.5 V is applied to the electrodes. SC electrodes are usually attached to the medial or distal phalanx of the index and middle finger, usually of the non-dominant hand of the participant, so that they are able to use their dominant hand during (or before/after) the experimental procedure – for example, to fill in a short questionnaire measuring self-reported momentary emotional states.

In the following sections we discuss the use of SC in experimental research and the applicability of this research method to studies on translators and interpreters. Some methodological considerations are presented, focusing on standards for SC measurement. Potential problems related to the use of SC in CTIS are discussed and some solutions are suggested.

Rojo López, A. M., & Korpal, P. (2020). Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 191–217.

4.1. Skin conductance in translation and interpreting

In the past few decades, SC has been used to investigate emotions and stress experienced by translators and interpreters. Emotion is a theoretical construct. That is why in empirical research one looks for observable measures of emotion. SC responses and SC level have long been used to this end in Emotion Studies. The advantage of SC is that "it is the only autonomic psychophysiological variable that is not contaminated by parasympathetic activity" (Braithwaite et al., 2013, p. 3). It is also a non-invasive method and it does not take much time to prepare participants for an experimental session. The equipment is relatively inexpensive, especially when compared to electroencephalography (EEG), which has also been employed to investigate emotional arousal.

Another important advantage of SC as a research method is that it indexes sympathetic response that is not under conscious control. However, when making a decision to include SC in an experimental setup, also in CTIS, one must take into consideration all the confounding variables that may distort the signal and compromise data quality (see Section 4.4). The ecological validity of SC is limited, even more so when participants are asked to perform a cognitively taxing language task such as translation.

As for stress, since interpreting is commonly referred to as a psychologically taxing activity, several researchers have touched upon the question of physiological and psychological stress related to the profession of conference and community interpreting. In some of these studies, SC was used as a measure of stress. Physiological manifestations of stress experienced by interpreters have been investigated in the study by Kurz (2002) – see Section 3.1. In her experiment, Kurz measured physiological stress responses in media interpreting, as opposed to regular conference interpreting. As stated, apart from GSR, HR was used to operationalize stress. In her study, Kurz (2002) observed significant differences in SC level between on-site conference interpreting and live TV interpreting. However, SC level measures failed to discriminate between professional and student interpreters.

One year later, Kurz (2003) compared the level of stress experienced in an interpreting task between the two experimental groups, experts and novices. To this end, SC level was adopted again along with HR. Consistent with her hypothesis, Kurz (2003) observed higher HR values in the group of trainees. However, no statistically significant differences were observed for SC measures. Importantly, Kurz's (2003) experiment was a pilot study on two professionals and three interpreting novices. A large-scale study testing potential SC level differences between student and professional interpreters would have been needed to yield greater statistical power and, in turn, provide a more comprehensive insight into SC as a potential measure of stress in studies comparing the performance of professional interpreters and interpreting trainees.

In a study by Korpal and Jasielska (2019), SC was measured to test whether interpreters are affected by a speaker's emotions. A group of 20 professional interpreters was asked to

Rojo López, A. M., & Korpal, P. (2020). Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 191–217.

simultaneously interpret two speeches, a neutral one and an emotional one. The results point to interpreters' tending to emotionally converge with the speaker, as manifested not only by SC responses, but also by the Positive and Negative Affect Schedule scores (Watson et al., 1988). Therefore, SC was adopted in a within-subject study designed to look for differences in interpreters' physiological response to neutral and affect-laden speech.

SC has to date been used in only a few experiments in Interpreting Studies. However, one can observe a growing popularity of EDA in T&I, which should soon be reflected in a growing number of EDA studies published in T&I journals. HR and HRV have been applied more often, in studies on both translators and interpreters (see Section 3.1). Although SC has been used for decades in Emotion Studies as a measure of physiological arousal, its use might significantly compromise the ecological validity of a study in which translation or interpreting is performed. In order not to distort the GSR signal, in studies using SC, participants are asked not to move the hand with SC electrodes attached to it. Even though electrodes attached to participants' fingers seem not to interfere significantly with performing simultaneous interpreting, the same cannot be said of translation.

This is probably the most obvious reason why SC appears not to have been used in a study tracking the process of translation. A solution might be to use either toes or the volar side of the wrist for electrodermal recording in T&I studies. However, when using a wristwatch with the EDA recording system, it should be considered that this site appears to "reflect thermoregulatory rather more than psychophysiologically relevant electrodermal phenomena" (Boucsein et al., 2012, p. 1022). In general, compromised ecological validity might be one of the main drawbacks of SC applied to studying translators and interpreters. Other limitations of using SC in Translation and Interpreting Studies are discussed in more detail in Section 4.4.

4.2. Which SC variables should be analysed?

Two main types of data were collected in studies applying electrodermal activity, that is, skin conductance responses (SCR) and SC levels. SCRs can be defined as "phasic (i.e., short-term) responses to specific external stimuli" (Christopoulos et al., 2016, p. 398). This type of measure can be contrasted with SC level, that is, tonic responses, "the overall EDA (for instance, throughout a full experimental session)" (Christopoulos et al., 2016, p. 398). Figure 4 shows an example of SC recorded in a 30-second block.
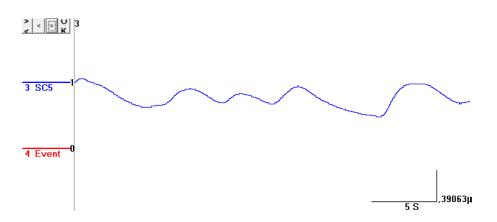
Rojo López, A. M., & Korpal, P. (2020). Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 191–217.

Figure 4. An example of recorded SC (PsychLab Analysis)

A few specific phasic responses (SCRs) can be identified in this 30-second block. The recorded signal can also be analysed in order to determine overall EDA measured in microsiemens (SC level – a tonic response).

To give an example, an experiment can be designed to test whether interpreting a given word or phrase triggers an emotional response in the participant. In such a study, it is important to measure the SCR latency, which is understood as the temporal distance between the stimulus presentation and the physiological response. The latency for SCR is usually between one and four seconds, although Christopoulos et al. (2016) report that responses longer than five seconds are also theoretically possible. In their SC manual, PsychLab suggests applying a time window of between one and four seconds for the onset of the SC response. A response observed later than that would be treated as a non-specific response or a spontaneous response (PsychLab, 2003). A classic SCR can be characterized by its onset, rising time, peak and exponential decay (PsychLab, 2003). SCRs are screened by their amplitude, that is, the difference between the onset and the peak of the response. In the past, 0.05 µS was adopted as a common threshold. However, with the advent of more precise technologies, minimum limits of 0.04–0.01 µS are now more often adopted in SC studies (Braithwaite et al. 2013).

Apart from SCRs, researchers may compare the overall SC level for each experimental condition. For example, in CTIS, general electrodermal activity during the process of translating a highly specialized text could be compared to that of a non-specialized text. In such a study, conclusions would be drawn on the basis of the general sympathetic activity manifested in a more pronounced SC level. However, averaging across the whole EDA signal may be an inadequate measure of SC level; it is "likely to over-estimate the true-SCL as such averages will also contain SCRs (thus artificially elevating the measure)" (Braithwaite et al., 2013, p. 5). One of the ways to solve this problem is to specify a short time period before each SC response; this may be considered a valid measure of the tonic SC level which does not include, or is not influenced by, a specific SCR (Braithwaite et al. 2013).

Rojo López, A. M., & Korpal, P. (2020). Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 191–217.

4.3. Should SC be treated as a dependent or an independent variable?

Similarly to HR and HRV, SC is usually the measure of a dependent variable in experimental research. In reception studies, it is often used to test whether affect-laden stimuli, such as pictures (Gatti et al., 2018), single words (Harris et al., 2003) or short narratives (Jankowiak & Korpal, 2018) evoke a pronounced emotional response. In psycholinguistics, the SC method has been used, for instance, to test the hypothesis whether participants' emotional reaction to stimuli presented in their non-native language would be attenuated when compared to the native tongue (Harris, 2004; Caldwell-Harris & Ayçiçeği-Dinn, 2009). Therefore, the effect of the language (independent variable) on physiological arousal manifested by SC (dependent variable) is tested in such an experimental design.

However, it is possible to think of a CTIS experimental design where SC would be a measure of an independent variable. In such a study, an increase in SC level might be used as an operational definition of the psychological stress experienced by a translator or an interpreter:

- Does stress have an impact on translators' performance?
- Does stress increase the number of disfluencies in interpreters' output?
- Can stress increase the number of mistakes made by interpreters while performing sight translation?

These are a few potential research questions in which SC could be applied as the measure of an independent variable.

4.4. Which confounding variables should be controlled?

Researchers using SC as a research method in CTIS should realize that the experimental setup might intimidate the participants in a study. Participants might even be apprehensive of the fact that the physiological reactions of their bodies are measured while they perform a translation or an interpreting task. Another issue is that sweat gland activity might result not only from participants' emotional excitement, but also from stress related to their performance being evaluated (cf. Korpal & Jankowiak, 2018). That is why it seems to be of the utmost importance to follow strictly all the regulations and guidelines regarding the ethical use of human subjects in research. At the beginning of the experiment, participants should have enough time to relax and familiarize themselves with the experimental set-up. Before the experiment, when the participant rests, a baseline SC can be recorded (cf. Gatti et al., 2018). Not providing enough time for participants to read the instructions and relax before the task might distort the baseline SC value. Importantly, participants' physiological arousal during the experiment might result from anxiety and not from emotional excitement in response to an affect-laden stimulus. Another methodological issue is that SC does index physiological arousal but not necessarily the valence of a given emotion. Although the literature distinguishes between high-arousal and low-arousal emotions (e.g., Kensinger,

2004; Kensinger & Corkin, 2004), applying SC as the only method of investigating emotion might not be sufficient to answer the question of which emotion, and of what valence, the participant experiences at a given point in time as a result of experimental manipulation.

There are other confounding variables. SC is believed to be influenced by ambient temperature and humidity. To obtain valid results, all participants should be recorded in similar conditions. For example, Christopoulos et al. (2016) suggest that around 25 °C to 26 °C with 50 per cent humidity is a comfortable temperature, whereas Braithwaite et al. (2013) recommend that an ambient room temperature of 22–24 °C be maintained.

Any medication taken by study participants may influence SC recording. This is of crucial importance in longitudinal studies where the same participant is tested again after some time. In CTIS, the effect of training on stress experienced by translators or interpreters might be tested in a SC study. To minimize the impact of medication as a confounding variable, information on participants' medication and health condition should be collected in a self-report questionnaire administered either before or after the experiment.

What is more, individual differences and life experiences may have an impact on emotional excitement. Sweating has also been proved to be gender- and age-related (Christopoulos et al., 2016; Morimoto, 1978). Electrodermal activity may also vary with respect to race and body mass index (BMI) (Doberenz et al., 2011). Other factors that may influence SC measurement are smoking (Furedy et al., 1999), physical properties of the skin (Allen et al., 1973; Doberenz et al., 2011) and caffeine intake (Barry et al., 2005). Physical activity may also possibly influence SC measurement, but the existing research appears to be inconclusive on this matter. Some individuals may also be hypo-responsive regarding their EDA and their data might need to be discarded (Braithwaite et al., 2013, p. 41). For a detailed report on potential confounding variables in SC, see Doberenz et al. (2011).

To summarize, many methodological challenges of SC studies are caused by significant intra-group variation (see also Korpal & Jankowiak, 2018 for the methodological considerations related to physiological measures applied in empirical studies). Collecting data from a larger number of participants might help to minimize the effect of individual differences on results.

Finally, apart from individual differences, habituation may be a problem to solve in studies using SC (Christopoulos et al., 2016; Gatti et al., 2018). As explained by Gatti et al. (2018)

> the effect of the triggering event on the SC signal might also decrease after several presentations of emotional stimuli, as the participant could habituate to the emotional stimulation itself, … a participant could grow accustomed to the emotional impact of the eliciting stimuli. (pp. 9–10)

Counterbalanced stimuli presentation might at least partly mitigate this effect.

Rojo López, A. M., & Korpal, P. (2020). Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 191–217.

4.5. How many participants are needed?

Increasing the external validity of the study by collecting data from more participants is important in studies applying physiological measurement because of significant intra-group variation. As observed by Kurz (2003), who conducted an SC study on interpreters, "it seems that SCL is better suited to measure intra-individual differences" (p. 62), as opposed to differences between participants and experimental groups. It appears that at least 20 participants are needed to be able to discriminate between different types of stimuli used in a study (Christopoulos et al., 2016). As the population of translators and interpreters is relatively small, a within-subject design, where each participant is tested in at least two conditions, might be a better choice than a between-subject design for a T&I SC study.

4.6. Which experiment structure is recommended?

The specific structure of an SC experiment depends on the type of data, that is, whether the focus is on tonic (overall EDA) or phasic (short-term) responses. Irrespective of the type of responses, however, researchers should leave enough time before the experiment (a few minutes) for the electrode gel to be absorbed (Braithwaite et al., 2013). Also, they should run a baseline measurement in every SC experiment. At this time, the participants should not be engaged in any task. If the experiment involves the presentation of stimuli on slides, the participants may be asked to look at a fixation cross displayed on the computer screen. Braithwaite et al. (2013) recommend using a 2–4-minute baseline session. They also suggest that a one-minute baseline period also be used in any experimental signal (e.g., a speech/story), preferably before the measurement period (Braithwaite et al., 2013). In terms of the experiment structure for event-related SCRs, it is crucial to consider the SCR latency – up to four seconds. Failure to adopt a recommended time frame may make it impossible to pair SCRs to specific stimuli presented to participants as part of the experimental procedure.

4.7. Other standards and recommendations for SC measurement and data analysis

Some recommendations related to SC measurement have been mentioned, including the difference between SCR and SC level, SCR latency and a minimum amplitude for measured SC. The specific software and hardware used in the study may have specific guidelines attached to them. For example, for the BIOPAC MP36R System and an experiment lasting approximately 60 minutes, Braithwaite et al. (2013) suggest the following settings: EDA channel sample rate, 1–2 KHz; acquisition sampling rate, 2000–5000 samples/sec; gain × 2000 (however, gain × 1000 is sufficient in most cases).

The analysis of the signal usually starts with a visual inspection. Some artefacts may be detected at this early stage of data analysis: for example, "periods of poor contact or sharp square-wave spiking that may reflect contamination from an artificial source" (Braithwaite et al., 2013, p. 15). The selection of the appropriate data-analysis method depends to a large extent on the software used. In PsychLab Analysis software, in order to analyse collected SC data, the observed signal is sectioned into time blocks and then processed to (1) estimate SC

Rojo López, A. M., & Korpal, P. (2020). Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 191–217.

level and (2) detect SCRs in each moment of interest. The signal is down-sampled so that data can be processed faster. Any movement artefacts should be identified and then corrected (Gatti et al., 2018); artefacts can be removed by down-sampling the signal. On some occasions, when the artefacts are too excessive, some data have to be discarded (Braithwaite et al., 2013). After proper correction, the signal can be analysed for the number of SCRs and SC levels. (For a detailed description of the data-analysis process for the BIOPAC MP36R System, refer to Braithwaite et al., 2013).

5. A final remark on ethics and triangulation

As a final remark concerning HR/HRV/SC measurement, two notions should be discussed: ethics and triangulation. As for ethical issues, the procedure to fit both HR/HRV and SC devices might be rather invasive since it involves touching participants. In the case of HRV, the procedure might be particularly intrusive, since fitting a chest band requires raising the participant's clothes just below their chest. To minimize discomfort, the participant should be allowed to choose (the gender of) the person providing the assistance. It is also important to describe the whole procedure in advance to the participants and explain to them that using SC/HR/HRV is by no means detrimental to their health.

Moreover, investigating topics such as emotional regulation might be potentially invasive, and participants might not be willing to share their emotional states with researchers. This is likely to be the case especially when, apart from physiological methods, questionnaires are administered to participants, the experimental procedure possibly leading to personal self-disclosure (cf. Tourangeau & Yan, 2007). It is therefore important to indicate and ensure respect for data protection and privacy, explaining to the participants that the data will be collected, processed and stored in a secure and confidential way. It might also be worth explaining that all the stimuli have been designed for the sake of the study and that they do not refer to real events. As in any other study with human subjects, the participants should be told that they can withdraw from the experimental procedure at any time with no need to provide an explanation for their decision. They may also withdraw their consent to the processing of personal data at any time.

Triangulation is most generally understood as the use of multiple measures to capture a construct, but it can also be applied to various ways in which treatments and interventions may be operationalized, and even to the use of multiple theories, analyses, methodologies or research designs. Triangulation research techniques enable different dimensions of the same phenomenon to be explored and, even better, encourage cross-validation. The importance of research method triangulation cannot be over-emphasized. HR/HRV and SC index physiological arousal and, by and large, are less sensitive to emotional valence (negative/state of displeasure and positive/state of pleasure). To gain a better insight into emotion processing, therefore, HR/HRV and SC should be triangulated with other measures of emotions, including self-reports, such as the Positive and Negative Affect Schedule scores (PANAS; Watson et al., 1988), and interviews administered after the experiment.

Rojo López, A. M., & Korpal, P. (2020). Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 191–217.

6. Conclusions and emerging challenges

For a long time, researchers on emotion have striven to draw sharp boundaries between what were considered to be different components of emotions (Scherer, 1984) in order to encapsulate them into discrete categories with identifiable labels, such as *arousal*, *affect*, *feelings*, *mood* or *emotion*. But would such an attempt make as much sense when it is assumed – as postulated by the constructed theory of emotion – that they all result from the same biological mechanisms operated by brain and body in unison? Certainly, the case would be less clear-cut from a theoretical point of view, blurring as it does almost any distinction made in traditional theories (e.g., brain/body, cognition/emotion, emotion generation/regulation, to name just a few).

The case may not be as simple from a methodological perspective, though: research procedures have always been aided and guided by categorization, structuring and sequencing. From such a standpoint, the fact that an increased HR, body temperature and SC may not always be related to stress may be felt somehow as a methodological nuisance, but one certainly not as big as to throw researchers into despair. Rather, recent neuroscientific evidence should serve as a wake-up call, drawing researchers' attention to a pressing need – felt not only in psychophysiology but also in most research areas – for more comprehensive, in-depth protocols based on triangulation.

We would be closer to the finishing line if replicability and comparison across studies were encouraged by greater transparency, more detailed descriptions of procedures and a willingness to share raw and processed data. Obviously, this desideratum involves cooperation from editors and managers of journals, book series and publishing houses to circumvent space restrictions, finding convenient – and, if necessary, even more creative – ways to make information available. The time is now more right than ever for such a change, with some journals actually starting to require publication of all raw data and even making some space available on their websites to upload and display this information.

These practices are particularly desirable in areas such as HRV and SC studies, where so much variability is found on research measures, variables, instruments, and data analysis and reporting. We sincerely hope that this article will contribute to reducing the existing inconsistencies in how we experience and construct an instance of an emotion in a particular context, and that it will do so by finding a more direct and clearer way to relate what our heart and skin tell us about our physiological states. Now that neuroscience is helping to clear a few paths to the brain, a solid methodological headway is needed, and this article can help us take a significant step in the right direction by, both literally and figuratively, listening to our hearts and sensing through our skin.

Rojo López, A. M., & Korpal, P. (2020). Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 191–217.

References

AIIC (International Association of Conference Interpreters). (2002). *Interpreter workload study: Full report*. http://aiic.net/page/657/interpreter-workload-study-full-report/lang/1.

Allen, J. A., Armstrong, J. E., & Roddie, I. C. (1973). The regional distribution of sweating in man. *Journal of Physiology, 235*(3), 749–759. https://doi.org/10.1113/jphysiol.1973.sp010414

Barrett, L. F. (2017a). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12*(1), 1–23. https://doi.org/10.1093/scan/nsx060.

Barrett, L. F. (2017b). *How emotions are made: The secret life of the brain*. Houghton Mifflin Hardcourt.

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10*(1), 20–46. https://doi.org/10.1207/s15327957pspr1001_2

Barry, R. J., Rushby, J. A., Wallace, M. J., Clarke, A. R., Johnstone, S. J., & Zlojutro I. (2005). Caffeine effects on resting-state arousal. *Clinical Neurophysiology, 116*(11), 2693–2700. https://doi.org/10.1016/j.clinph.2005.08.008

Bassett, D. (2016). A literature review of heart rate variability in depressive and bipolar disorders. *Australian and New Zealand Journal of Psychiatry, 50*(6), 511–519. https://doi.org/10.1177/0004867415622689

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology, 49*(8), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1990). Startle reflex modification: Emotion or attention? *Psychophysiology, 27*(5), 513–522. https://doi.org/10.1111/j.1469-8986.1990.tb01966.x

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). *A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments*. https://www.birmingham.ac.uk/Documents/college-les/psych/saal/guide-electrodermal-activity.pdf

Caldwell-Harris, C. L., & Ayçiçeği-Dinn, A. (2009). Emotion and lying in a non-native language. *International Journal of Psychophysiology, 71*(3), 193–204. https://doi.org/10.1016/j.ijpsycho.2008.09.006

Cannon, W. B., & Rosenberg, C. E. (1932). *Wisdom of the body*. W.W. Norton and Company.

Cook, E. W., Hawk, L. W., Davis, T. L., & Stevenson, V. E. (1991). Affective individual differences and startle reflex modulation. *Journal of Abnormal Psychology, 100*(1), 5–13. https://doi.org/10.1037/0021-843X.100.1.5

Chernigovskaya, T. V., Parina, I. S., Alekseeva, S. V., Konina, A. A., Urich, D. K., Mansurova, Yu. O., & Parin, S. B. (2019). Simultaneous interpreting: Characteristic of autonomic provision of extreme cognitive loads. *Sovremennye Tehnologii v Medicine, 11*(1), 132–140. https://doi.org/10.17691/stm2019.11.1.15

Christopoulos, G. I., Uy, M. A., & Yap, W. J. (2016). The body and the brain: Measuring skin conductance responses to understand the emotional experience. *Organizational Research Methods*, *22*(1), 394–420. https://doi.org/10.1177/1094428116681073

Rojo López, A. M., & Korpal, P. (2020). Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 191–217.

Davis, M. (1992). The role of the amygdala in conditioned fear. In J. P. Aggleton (Ed.), *The amygdala: Neurobiological aspects of emotion, memory and mental dysfunction* (pp. 255–306). Wiley-Liss.

Doberenz, S., Roth, W. T., Wollburg, E., Maslowski, N. I., & Kim, S. (2011). Methodological considerations in ambulatory skin conductance monitoring. *International Journal of Psychophysiology, 80*(2), 87–95. https://doi.org/10.1016/j.ijpsycho.2011.02.002

Ekman P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Frijda, N. H., & Scherer, K. R. (2009). Emotion definition (psychological perspectives). In D. Sander & K. R. Scherer (Eds.), *Oxford companion to emotion and the affective sciences* (pp. 142–143). Oxford University Press.

Furedy, J. J., Algan, O., Vincent, A., Demirgoren, S., & Pogun, S. (1999). Sexually dimorphic effect of an acute smoking manipulation on skin resistance but not on heart-rate during a cognitive verbal task. *Integrative Physiological and Behavioral Science, 34*, 219–226. https://doi.org/10.1007/BF02688690

Gatti, E., Calzolari, E., Maggioni E., & Obrist, M. (2018). Emotional ratings and skin conductance response to visual, auditory and haptic stimuli. *Scientific Data, 5*, 1–12. https://doi.org/10.1038/sdata.2018.120

Grabner, R. H., Brunner, C., Leeb, R., Neuper, C., & Pfurtscheller, G. (2007). Event-related EEG theta and alpha band oscillatory responses during language translation. *Brain Research Bulletin, 72*(1), 57–65. https://doi.org/10.1016/j.brainresbull.2007.01.001

Halverson, S. L. (2010). Cognitive Translation Studies: Developments in theory and method. In G. Shreve & E. Angelone (Eds.), *Translation and cognition* (pp. 349–369). John Benjamins. https://doi.org/10.1075/ata.xv.18hal

Harris, C. L. (2004). Bilingual speakers in the lab: Psychophysiological measures of emotional reactivity. *Journal of Multilingual and Multicultural Development, 25*(2–3), 223–247. https://doi.org/10.1080/01434630408666530

Harris, C. L., Ayçiçeği, A., & Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first than in a second language. *Applied Psycholinguistics, 24*(4), 561–579. https://doi.org/10.1017/S0142716403000286

Heathers, J. A. J. (2014). Everything Hertz: Methodological issues in short-term frequency-domain HRV. *Frontiers in Psychology, 5*(177), 1–15. https://doi.org/10.3389/fphys.2014.00177

Hervais-Adelman, A., Moser-Mercer, B., & Golestani, N. (2015). Brain functional plasticity associated with the emergence of expertise in extreme language control. *NeuroImage, 114*, 264–274. https://doi.org/10.1016/j.neuroimage.2015.03.072

Hvelplund, K. T. (2014). Eye-tracking and the translation process: Reflections on the analysis and interpretation of eye-tracking data. *Minding Translation, Special issue 1,* 201–223. https://doi.org/10.6035/MonTI.2014.ne1.6

Iturregui-Gallardo, G., Serrano-Ratera, A., Méndez-Ulrich, J.-L., Soler-Vilageliu, O., & Jankowska, A. (2018, June 4–5). *Testing the filmic experience: Audio subtitling and psychophysiology* [Paper presentation]. Understanding Quality in Media Accessibility, Barcelona, Spain.

Jankowiak, K., & Korpal, P. (2018). On modality effects in bilingual emotional language processing: Evidence from galvanic skin response. *Journal of Psycholinguistic Research, 47*(3), 663–667. https://doi.org/10.1007/s10936-017-9552-5

Kensinger, E. A. (2004). Remembering emotional experiences: The contribution of valence and arousal. *Reviews in the Neurosciences, 15*(4), 241–251. https://doi.org/10.1515/REVNEURO.2004.15.4.241

Rojo López, A. M., & Korpal, P. (2020). Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 191–217.

Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences, 101*(9), 3310–3315. https://doi.org/10.1073/pnas.0306408101

Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H., & Koo, B.-H. (2018). Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry Investigation, 15*(13), 235–245. https://doi.org/10.30773/pi.2017.08.17

Kleckner, I. R., Zhang, J., Touroutoglou, A., Chanes, L., Xia, C., Simmons, W. K., Quigley, K. S., Dickerson, B. C., & Barrettt, L. F. (2017). Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nature Human Behaviour, 1*, Article 0069, 1–14. https://doi.org/10.1038/s41562-017-0069

Klonowicz, T. (1994). Putting one's heart into simultaneous interpretation. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation* (pp. 213–224). John Benjamins. https://doi.org/10.1075/btl.3.16klo

Korpal, P. (2016). Interpreting as a stressful activity: Physiological measures of stress in simultaneous interpreting. *Poznań Studies in Contemporary Linguistics, 52*(2), 297–316. https://doi.org/10.1515/psicl-2016-0011

Korpal, P. (2017). *Linguistic and psychological indicators of stress in simultaneous interpreting*. Wydawnictwo Naukowe UAM.

Korpal, P., & Jankowiak, K. (2018). Physiological and self-report measures in emotion studies: Methodological considerations. *Polish Psychological Bulletin, 49*(4), 475–481.

Korpal, P., & Jasielska, A. (2019). Investigating interpreters' empathy: Are emotions in simultaneous interpreting contagious? *Target, 31*(1), 2–24. https://doi.org/10.1075/target.17123.kor

Kurz, I. (2002). Physiological stress responses during media and conference interpreting. In G. Garzone & M. Viezzi (Eds.), *Interpreting in the 21st century: Challenges and opportunities* (pp. 195–202). John Benjamins. https://doi.org/10.1075/btl.43.19kur

Kurz, I. (2003). Physiological stress during simultaneous interpreting: A comparison of experts and novices. *The Interpreters' Newsletter, 12*, 51–67.

Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research: Recommendations for experiment planning, data analysis, and data reporting. *Frontiers in Psychology, 8*, 1–18. https://doi.org/10.3389/fpsyg.2017.00213

LeDoux, J. E. (1991). Emotion and the limbic system concept. *Concepts in Neuroscience, 2*, 169–199.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett. L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences, 35*(3), 121–143. https://doi.org/10.1017/S0140525X11000446

Malik, M. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal, 17*(3), 354–381. https://doi.org/10.1093/oxfordjournals.eurheartj.a014868

Melillo, P., Bracale, M., & Pecchia, L. (2011). Nonlinear heart rate variability features for real-life stress detection. Case study: Students under stress due to university examination. *BioMedical Engineering OnLine, 10*(96). https://doi.org/10.1186/1475-925X-10-96

Morimoto, T. (1978). Variations of sweating activity due to sex, age and race. *Physiology and Pathophysiology of the Skin, 5*, 1655–1666.

Muñoz Martín, R. (2010). On paradigms and cognitive translatology. In G. M. Shreve & E. Angelone (Eds.), *Translation and cognition* (pp. 169–187). John Benjamins. https://doi.org/10.1075/ata.xv.10mun

Rojo López, A. M., & Korpal, P. (2020). Through your skin to your heart and brain: A critical evaluation of physiological methods in Cognitive Translation and Interpreting Studies. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 191–217.

Orero, P., Doherty, S., Kruger, J.-L., Matamala, A., Pedersen, J., Perego, E., Romero-Fresco, Rovira-Esteva, S., P., Soler-Vilageliu, O., & Szarkowska, A. (2018). Conducting experimental research in audiovisual translation (AVT): A position paper. *Journal of Specialised Translation, 30*, 105–126.

Pace-Schott, E. F., Amole, M. C., Aue, T., Balconi, M., Bylsma, L. M., Critchley, H., Demaree, H. A., Friedman, B. H., Kotynski Gooding, A. E., Gosseries, O., Jovanovic, T., Kirby, L. A. J., Kozlowska, K., Laureys, S., Lowe, L., Magee, K., Marin, M.-F., Merner, A. R., Robinson, J. L., Smith, R. C., Spangler, D. P., Van Overveld, M., & VanElzakker, M. B. (2019). Physiological feelings. *Neuroscience & Biobehavioral Reviews, 103*, 267–304. https://doi.org/10.1016/j.neubiorev.2019.05.002

Penachini da Costa de Rezende Barbosa, M., Turri da Silva, N., Mícolis de Azevedo, F., Marcelo Pastre, C., & Marques Vanderlei, L. C. (2014). Comparison of Polar® RS800G3™ heart rate monitor with Polar® S810i™ and electrocardiogram to obtain the series of RR intervals and analysis of heart rate variability at rest. *Clinical Physiology and Functional Imaging, 36*(2), 112–117. https://doi.org/10.1111/cpf.12203

Piskorski, J., & Guzik, P. (2005). Filtering Poincaré plots. *Computational Methods in Science and Technology, 11*(1), 39–48. https://doi.org/10.12921/cmst.2005.11.01.39-48

PsychLab. (2003). *Analysis for recordings in psychophysiology: Software manual*. http://www.PsychLab.com/download.html.

Quintana, D. S. (2016). Statistical considerations for reporting and planning heart rate variability case-control studies. *Psychophysiology, 54*(3), 344–349. https://doi.org/10.1111/psyp.12798

Ramos Caro, M. (2015). The emotional experience of films: Does Audio Description make a difference? *The Translator, 21*(1), 68–94. https://doi.org/10.1080/13556509.2014.994853

Ramos Caro, M. (2016). *La traducción de los sentidos : Audiodescripción y emociones*. LINCOM.

Rojo López, A. M., & Ramos Caro, M. (2014). The impact of translators' ideology on the translation process: A reaction time experiment. In R. Muñoz Martín (Ed.), *Minding Translation*. *MonTi*, Special Issue 1, 247–272.

Rojo López, A. M., Ramos Caro, M., & Valenzuela Manzanares, J. (2014). The emotional impact of translation: A heart rate study. *Journal of Pragmatics, 71*, 31–44. https://doi.org/10.1016/j.pragma.2014.07.006

Roziner, I., & Shlesinger, M. (2010). Much ado about something remote: Stress and performance in remote interpreting. *Interpreting, 12*(2), 214–247. https://doi.org/10.1075/intp.12.2.05roz

Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293–317). Psychology Press.

Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory (from Y1 – Y2)*. Consulting Psychologists Press.

Szpak, K. S., Alves, F., & Buchweitz, A. (in press). Perspective taking in translation: In search of neural correlates of representing and attributing mental states to others. In R. Muñoz Martín & S. L. Halverson (Eds.), *Multilingual mediated communication and cognition* (pp. 133–154). Routledge. https://doi.org/10.4324/9780429323867-6

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859–883. https://doi.org/10.1037/0033-2909.133.5.859

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Waugh, C. E., Thompson, R. J., & Gotlib, I. H. (2011). Flexible emotional responsiveness in trait resilience. *Emotion, 11*(5), 1059–1067. https://doi.org/10.1037/a0021786