Physiological Versus Self-Report Measures of Arousal During Tactical Training Involving a Synthetic Topographic Environment

Michael W. Boyce, Ph.D.
Combat Capabilities and Development Center - Soldier Center (CCDC-SC)
Simulation and Training Technology Center (STTC)
12423 Research Parkway, Orlando, FL 32826
michael.boyce11.civ@mail.mil

Cortnee Stainrod
University of Central Florida – Institute for Simulation & Training
3100 Technology Pkwy, Orlando, FL 32826
cortnee.stainrod.ctr@mail.mil

Aryn Pyke, Ph.D.
Army Cyber Institute, United States Military Academy
West Point, NY
aryn.pyke@westpoint.edu

Keywords:
SAM, Self-Assessment Manikin, EDA, ARES

ABSTRACT: *This research examines the relationship between the electrodermal activity (EDA) of 43 West Point Cadets while viewing military tactics displays and compares that to results from the Self-Assessment Manikin (SAM; Bradley & Lang, 1994). First there is a need to understand how EDA varies between two different types of presentation formats. Second it was expected that the EDA data would negatively correlate to self-report data based on previous research (Boyce, Reyes, et al., 2016), and third was that EDA and self-report data would be able to predict performance. Results did not indicate significant differences based on display type, however the results did support the negative correlation between EDA and SAM. Finally there was a trend toward predicting performance but it did not reach statistically significant levels, warranting the need for further investigation.*

1. Introduction

With the modernization of the Army focusing on rapidly equipping soldiers through modeling and simulation (M&S), there is a need to understand which systems are appropriate when and for whom. To make informed decisions, data can be collected using a variety of techniques to include self-report assessments, performance measures and physiological responses. Self-report assessments provide insight on how an individual may think or feel throughout an experience. Physiological measures are those which can detect signals coming directly from the body. Collecting with any one type of measure can lead to an incomplete picture, which leads researchers to cross-validate using multiple techniques [2]. This paper discusses the differences between self-report scores (using the Self-Assessment Manikin) and physiological measurements (using electrodermal activity) for military tactical decision-making on the ARES Sandtable.

2. Background

The Self-Assessment Manikin (SAM) is a picture-oriented scale to assess the affect dimensions of valence (pleasure), arousal, and dominance. SAM is composed of three sets of five figures (manikins), which stand for the three dimensions [6]. The SAM is composed of 3 sets of 5 figures (manikins), which stand for the 3 major affective dimensions (see Figure 1). The SAM has been evaluated across population types, genders, and groups of students such as military cadets [7] [8] [10].

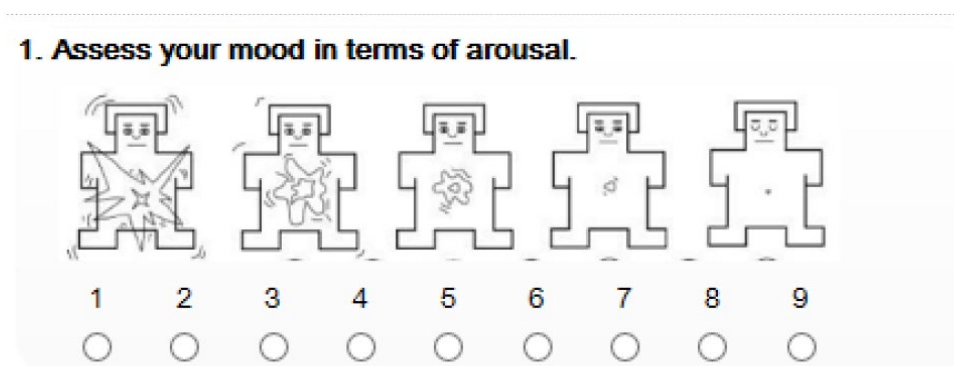

Figure 1: Example of Self-Assessment Manikin Prompt

Electrodermal Activity (EDA), also known as Galvanic Skin Response, is a measure of electrical current coming from the skin and can serve as an indicator of physiological change within a research participant. Past research connects EDA to arousal, engagement, and stress [9] [11]. Perala and Sterling (2007) recommend using EDA-type solutions due to their noninvasiveness, reliability, and ability to collect data throughout soldier training when measuring stress. In their report they do recommend using multiple methods, since although EDA can detect a change in emotion, identifying the specific emotion is not possible such as stress, anxiety, or excitement. As an example of this, Schaefer and colleagues (2019) using combinations of wearable EDA sensors and self-report data were able to distinguish frustration from excitement or sensation seeking from a team of Marines in a combat gunnery task as a way of measuring stress. The takeaway message is that EDA can be used as part of a larger set of measurements providing comparison against other metrics which is relevant to this research.

The Augmented REality Sandtable (ARES) [1] [4] [5] is a traditional sand table, filled with play sand, augmented with a commercial off-the-shelf (COTS) projector, LCD monitor, laptop, and Microsoft Kinect and Xbox Controllers. ARES is an example of a planar display on which the participant will view images of a military tactics scenario via central perspective projection. Central perspective involves projecting directly, without tilt, onto a surface below, providing depth cues of the terrain. For this experiment, the ARES projection technology was combined with terrain boards rather than the actual sand table.

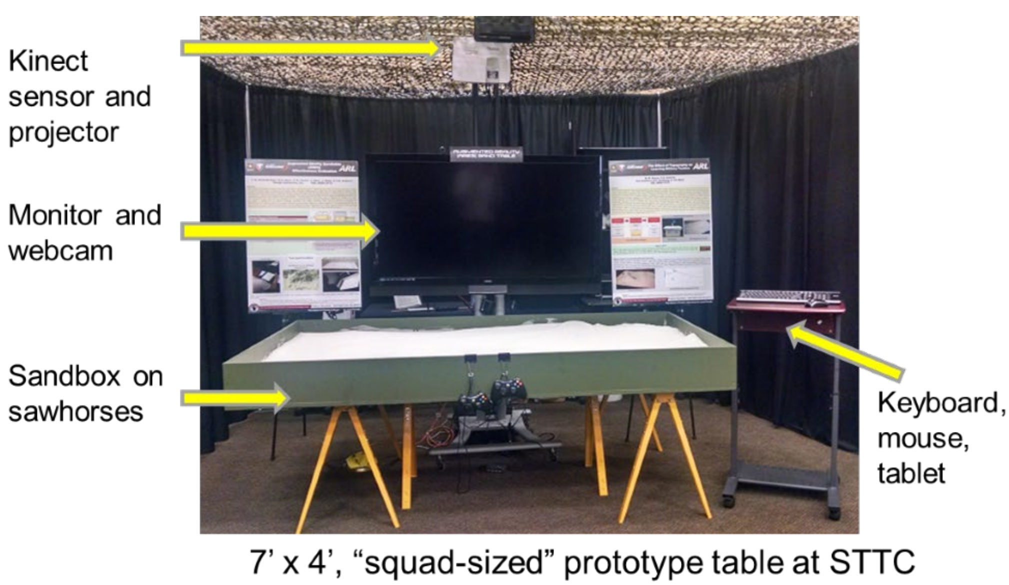

Figure 2: ARES Sandtable

This experiment is an expansion of Boyce, Goldberg, and Moss (2016) who ran a study with 19 ROTC cadets looking for differences in electrodermal activity / skin conductance responses between flat and raised surface types (see Figure 3). Although the results were too small to achieve statistical significance, data showed evidence that individuals using the flat surface exhibited more skin conductance responses. This data was in the opposite direction of Self-Assessment Manikin [6] scores as reported in Boyce, Reyes and colleagues (2016), which demonstrated that individuals who used the raised surface reported an increase in excitement (as mentioned by the arousal subscale) and an increased feeling of being in control of the tactical situation as opposed to being out of control (measured by the dominance subscale). Therefore, there was a desire to do a follow-on study with a larger population to deconflict results.

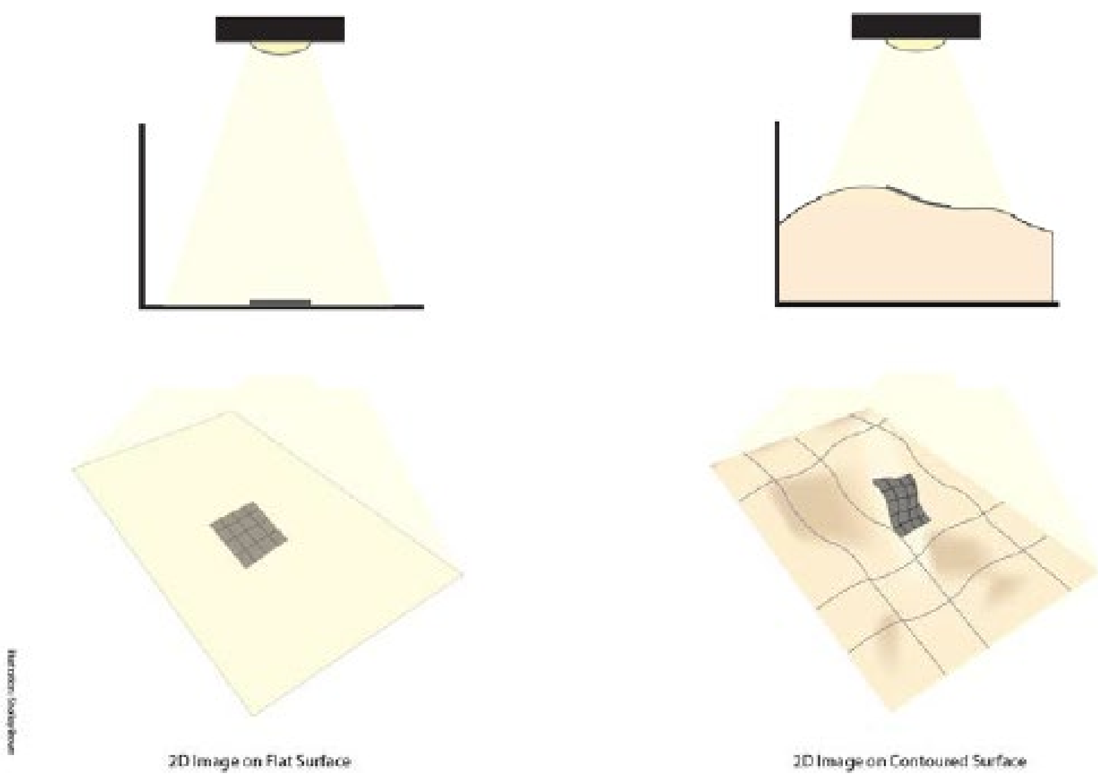

Figure 3: ARES Surface Differences

2.1 Research Questions and Hypotheses

Research Question 1: Is there a significant effect of EDA levels based on surface type?

H_1: There will be a significant effect of EDA level when controlling for baseline such that individuals in the flat condition will have higher levels than individuals in the raised condition.

Research Question 2: Is there a significant effect of SAM subscales based on surface type?

H_2: There will be a significant correlation between of the SAM and EDA when controlling for baseline when collapsed across conditions.

Research Question 3: EDA and the SAM predict performance?

H_3: EDA and SAM will be significant predictors on performance for cadets learning military tactics tasks.

3. Method

3.1 Participants

A total of 62 cadets from the United States Military Academy (USMA) participated in the study, Participants were from introductory psychology courses via USMA Sona System (M_{age} = 18.44, SD_{age} = .80; 28 males and 15 females). Cadets received extra credit in their classes for their participation of the study.

3.2 Equipment

Microsoft Band 2

The physiological variable of EDA was monitored to analyze physiological response associated with arousal during execution of tasks. The sensor used for the study is the Microsoft Band 2. The Microsoft Band 2 is a wearable, wireless biosensor that measures emotional arousal via skin conductance, a form of EDA that grows higher during states such as excitement, attention, or anxiety and lower during states of boredom or relaxation. The sensor also measures temperature and movement. The device will be used to monitor participants' arousal response to the training materials and interactions experienced during the study.

3.3 Procedures

Prior to arrival, participants were randomized and counterbalanced into one ordering of the two conditions, with either the flat condition or the raised condition presented first, followed by the opposite condition. Upon arrival, participants received a brief overview of the study and were asked to complete an informed consent form. Next, participants answered a demographics survey and Self-Assessment Manikin (SAM) to establish pre-experimental/baseline levels of subjective arousal and affect. Following this, participants were fitted with the Microsoft Band 2.

Next, participants were given a short introduction explaining the scenario, specifically that they were going to be asked squad and platoon level tactics questions and explaining the concepts to be covered in the lesson. Changes in EDA were expected during this task due to the presence or absence of additional depth information which could change their affective response based on engagement, perception of ability to answer the questions with the current surface or general pleasure. This introduction was followed by a series of practice examples where the participant received questions, answered them and received feedback. It was during feedback that the participant was able to ask questions for clarification. It took approximately 40 minutes to perform all pre-experiment activities. Next, participants were placed in the one of the two experimental conditions (either using the flat surface or raised surface) which consisted of approximately 20 minutes of prompts followed by 15 minutes of survey questions for each condition. They then performed the same task using the other surface with questions of comparable content and difficulty. The participant completed the experiment in just under 2 hours.

3.4 Research Design

This experiment is a within-subjects design with surface type as the independent variable and dependent variables of EDA, performance data (accuracy), and SAM scores. The specific combinations depend on the analysis being conducted.

3.5 Data Preparation

The Microsoft Band provides data in resistance (measured in kilohms, kΩ), however in order to compare it with other methods used in previous works, a conversion to conductance is necessary. Conductance is measured in microsiemens, μS and the conversion formula is below (Boucsein, 2012). R in this formula equals the resistance value.

$$\mu S = 1000/(R[k\Omega]) \qquad (1)$$

Analysis focused on three time periods: A baseline measure, and a measure associated with each surface type, flat and raised. The baseline was established by taking the average EDA across the entire experiment. The EDA measure associated with each surface type was the average EDA while answering tactics questions on that surface type.

Although there are many possible ways to examine EDA data in these time windows, the average was chosen because it provided a single value which facilitated a straightforward comparison between the SAM and the EDA data. For SAM the baseline was taken before viewing either condition and then immediately after the participant finished answering tactics question on each surface type. The average values were taken and used to represent the data point.

Usable participant data were defined as any EDA file (each participant had one) that when looking across the entire length of the experiment the average EDA did not exceed 4 µS (high pass filter). The reasoning for this is that most skin conductance responses occur in the 2-3 µS range (Boucsein, 2012), and we had some participants that had extremely high EDA values, which likely indicated an equipment malfunction, high levels of movement, or other sources of error such as clothing covering part of the sensor. Out of the 62 potential participants, 17 exceeded the 4 µS threshold, and two had no data, bringing the total to 43 usable data points.

4. Results

4.1 Data Summary

Variable	Baseline		Flat Condition		Raised Condition	
	M	*SD*	*M*	*SD*	*M*	*SD*
SAM Arousal	5.63	1.70	5.67	1.89	5.26	1.99
SAM Dominance	5.42	1.01	5.28	1.55	5.88	1.40
SAM Pleasure	3.88	1.31	4.33	1.36	4.05	1.54
EDA	1.05	0.87	1.06	0.95	1.01	0.93
Accuracy	N/A	N/A	0.59	0.16	0.53	0.19

Table 1: Means and Standard Deviations for Variables of Interest

4.2 Hypothesis Testing

H_1: There will be a significant effect of EDA level when controlling for baseline such that individuals in the flat condition will have higher levels than individuals in the raised condition.

Results of a RMANOVA indicate a nonsignificant difference between conditions, $F(1,41) = .308$, $p = .582$. Therefore, H_1 is not supported.

H_2: There will be a significant correlation between of the SAM and EDA when controlling for baseline when collapsed across conditions.

Results of partial correlations indicate a significant negative correlation between EDA and the pleasure subscale, such that as pleasure increased, EDA decreased, $r = -.372$, $p = .020$. However, there were nonsignificant correlations on the arousal and dominance subscales, $r = -.049$, $p = .767$ and $r = -.063$, $p = .705$ respectively. Therefore, H_2 is partially supported.

H_3: EDA and SAM will be significant predictors on performance for cadets learning military tactics tasks.

Multiple regression analysis was used to test if EDA and SAM significantly predicted participants' accuracy when answering military tactics questions. The results of the regression indicated the two predictors explained 12% of the variance but was non-significant ($R^2 =.12$, $F(2,40) =2.62$, $p =.085$). It was found that the aggregate SAM score significantly predicted accuracy, ($\beta = .38$, $p =.041$), but EDA did not ($\beta = .030$, $p =.136$).

5. Discussion / Conclusion

This experiment seeks to answer the question of can mixed method designs be used in training experiments to better understand military training tasks. As more and more training occurs using wearable sensors, more data will be available from which researchers can analyze and interpret. Based on the findings here, it appears that the self-assessment manikin has some predictive capabilities. However, the EDA data seems to not have enough data / variance to make any definitive conclusions here.

One thought for the lack of significance on the EDA data is due to the high signal to noise ratio of the data. It could be that EDA is a good confirmatory measure for cross validation, but is not really well suited for not psychomotor / non stressful tasks. Also, it could be possible that with a more research grade sensor, a longer duration data collection, or a different set of analysis techniques that results could have been different. However as these types of wearable technologies continue to permeate the Army, testing and validation is both necessary and needed.

5. References

1. Amburn, C. R., Vey, N. L., Boyce, M. W., & Mize, J. R. (2015). *The augmented reality sandtable (ARES)*. Army Research Laboratory, Aberdeen Proving Ground, Aberdeen Maryland.
2. Azbel-Jackson, L., Butler, L. T., Ellis, J. A., & van Reekum, C. M. (2016). Stay calm! Regulating emotional responses by implementation intentions: Assessing the impact on physiological and subjective arousal. *Cognition and Emotion*, *30*(6), 1107–1121. https://doi.org/10.1080/02699931.2015.1049515
3. Boyce, M. W., Cruz, D., & Sottilare, R. (2017). *Interpretative Phenomenological Analysis for Military Tactics Instruction BT - Advances in Human Factors, Business Management, Training and Education* (J. I. Kantola, T. Barath, S. Nazir, & T. Andre, Eds.). Cham: Springer International Publishing.
4. Boyce, M. W., Reyes, R. J., Cruz, D. E., Amburn, C. R., Goldberg, B., Moss, J. D., & Sottilare, R. A. (2016). *Effect of Topography on Learning Military Tactics-Integration of Generalized Intelligent Framework for Tutoring (GIFT) and Augmented REality Sandtable (ARES), ARL-TR-7792*. Army Research Laboratory, Aberdeen Proving Ground, Aberdeen Maryland.
5. Boyce, M. W., Rowan, C. P., Shorter, P. L., Moss, J. D., Amburn, C. R., Garneau, C. J., & Sottilare, R. A. (2019). The impact of surface projection on military tactics comprehension. *Military Psychology*, *31*(1), 45–59. https://doi.org/10.1080/08995605.2018.1529487
6. 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.
7. DeFalco, J. A., Rowe, J. P., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B. W., … Lester, J. C. (2018). Detecting and addressing frustration in a serious game for military training. *International Journal of Artificial Intelligence in Education*, *28*(2), 152–193.
8. Geethanjali, B., Adalarasu, K., Hemapraba, A., Pravin Kumar, S., & Rajasekeran, R. (2017). Emotion analysis using SAM (Self-Assessment Manikin) scale. *Biomedical Research (0970-938X)*, *28*.
9. Hernandez, J., Morris, R. R., & Picard, R. W. (2011). *Call Center Stress Recognition with Person-Specific Models BT - Affective Computing and Intelligent Interaction* (S. D'Mello, A. Graesser, B. Schuller, & J.-C. Martin, Eds.). Berlin, Heidelberg: Springer Berlin Heidelberg.
10. Paquette, L., Rowe, J., Baker, R., Mott, B., Lester, J., DeFalco, J., … Georgoulas, V. (2016). Sensor-Free or Sensor-Full: A Comparison of Data Modalities in Multi-Channel Affect Detection. *International Educational Data Mining Society*.
11. Picard, R. W., Fedor, S., & Ayzenberg, Y. (2015). Multiple Arousal Theory and Daily-Life Electrodermal Activity Asymmetry. *Emotion Review*, *8*(1), 62–75. https://doi.org/10.1177/1754073914565517
12. Perala, C. H., & Sterling, B. S. (2007). *Galvanic skin response as a measure of soldier stress*. Army Research Laboratory, Aberdeen Proving Ground, Aberdeen Maryland.
13. Schaefer, K., Baker, A., Brewer, R., Patton, D., Canady, J., & Metcalfe, J. (2019). Assessing multi-agent human-autonomy teams: US Army Robotic Wingman gunnery operations. In M. S. Islam & T. George (Eds.), *Micro- and Nanotechnology Sensors, Systems, and Applications XI* (p. 82). https://doi.org/10.1117/12.2519302