College Behavioral Finance

White Paper

BENTLEY UNIVERSITY | User Experience Center

May 21, 2020

Elizabeth Rosenzweig
Rachael Kelly
Tyler Roberts
Natalie Wadia
Kate Betteridge
Jacob Davidson
Desmond Fang

erosenzweig@bentley.edu
781.891.2032 | www.bentley.edu/uxc

Table of Contents

Study Overview

During the spring semester of 2020 Bentley's User Experience Center (UXC) ran a study to better understand how iMotions emotion-detection software compares to traditional means of emotional assessment- specifically Galvanic Skin Response detection (GSR) and Self-Reported Measurements SAM. This study was the continuation of previous work the UXC completed for Face of Finance.

Original Study Goals

The goal of the study was to better understand how iMotions captures user motivation by correlating facial expression data, GSR, and self-reported emotions. Study design, key findings, as well as the business case of using a combination of iMotions and SAM measurements in usability studies will be addressed in this white paper.

We also had several secondary goals. For instance, to learn if a second playthrough of the game results in a change of emotion. These findings will be further detailed in a forthcoming blog post available on the UXC website.

Finally, as a result of this study we were able to identify several best practices we can recommend to UX practitioners when designing studies that use a combination of biometric data and self-assessment. These findings are available in a case study available on the Bentley UXC website.

Research Study Design

Payback Game

We had participants play a personal finance game "PAYBACK" (www.TimeForPayback.com) which informed users about the realities and consequences of the choices they make in college while balancing their student loan debt and other responsibilities. The game PAYBACK was chosen specifically because of its relevance to the behavioral finance industry, which was the area of study the UXC had previously run.

Study Structure

The study consisted of 10 moderated usability sessions of 60 minutes conducted over four days. Biometric and self-reported emotion data was collected from 10 participants during the game PAYBACK (TimeForPayback.com), a college student finance simulation. The present study included two separate opportunities for participants to play the game and incorporate feedback (broken into Game 1 and Game 2). This was done to continue the study of finance behavior and choice. We chose to run two Behavioral Finance gamifications to better understand how the game's feedback affects performance, and whether students better understand the consequences of financial decision when they receive feedback.

Timing

Activity	**Allotted Time**
Intro/briefing	5 min
Background questions	5 min
Game 1: Payback game 1st playthrough	20 min
Midterm follow up questions (iMotions paused)	5 min
Game 2: Payback game 2nd playthrough	20 min
Wrap-up questions (iMotions paused)	5 min
Total time	**60 min**

Designing for Accurate Data Collection

After the first round of testing in the summer of 2019, we learned more about iMotions' capabilities and limitations. As a result of those findings, we made serval adjustments to our study design to ensure accurate data collection.

1. **We had the participants play the game twice.** We did this to mitigate variability in the learning process as participants learned from the game's feedback. The two games then yielded two different sets of data points from participants: 1) changes in game choices to understand if learning happened over the course of gameplay, and 2) if any changes during the second playthrough of the game that resulted in changes of emotion.
2. **We adjusted our normal moderation style to ensure the biometric data iMotions collected was accurate.** To collect biometric data accurately, participants should not be speaking at all while the measurements is being taken. We paused our recordings at set

points which happened in between each academic year, then we asked the participants verbally respond to SAM to set up a reference data point, thereby segmenting our recordings to ensure the data collected by iMotions is accurate and reflects the real emotion of participants.

3. **We combined iMotions technology with SAM questionnaire.** In order to capture and evaluate the emotions of participants more completely, researchers decided to use the SAM in addition to iMotions technology. Using data from both iMotions and SAM, researchers hoped to get a better sense of how iMotions data correlates with self-reported data.

Data Collection

We gathered emotional expression data at specific points of the game via iMotions as well as Galvanic Skin Response (GSR) via a Shimmer device. We also had participants self-report their emotions using the Self-Assessment Manikin Scale (SAM), see Figure 1 below.

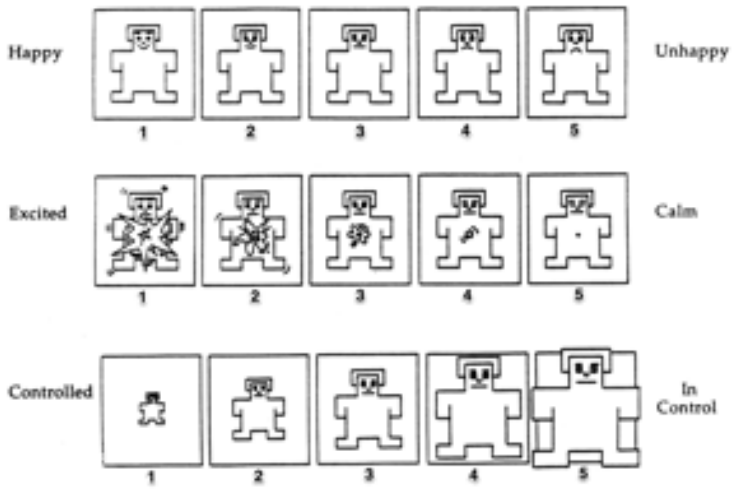

Figure 1 *Self-Assessment Manikin Scale (SAM)*

We included the self-report in order to see how the emotion detection matched up with the SAM and the GSR. Our analysis results are the triangulation between the SAM and biometric data and is discussed further in our findings.

Data Analysis

After the data had been collected, we analyzed the data both qualitatively and quantitatively. We performed statistical analysis of the biometric data collected in iMotions as well as the SAM averages across users and games. We did correlation between the averages of each participant's measured emotional response from iMotions and SAM as well as a t-test for the averages between both games. The results are detailed in our findings. We also qualitatively assessed verbal feedback that participants gave us at various points in the game and correlated this with the SAM scale and the biometric data. The findings from our analysis are detailed below.

Findings

Overview

The results of our study correspond with existing research which indicates emotions are difficult to measure even across people in a single situation (Barrett et al., 2019). In our data analysis, we discovered there was not always consistency between what a participant said, what they reported, and what iMotions measured.

Our key Quantitative findings are highlighted below:

- Participants showed more arousal in higher peaks/min as measured by GSR in Game 2. A t-test showed that this difference was significant.
- Participants reported a difference in happiness (valence) between Game 1 and Game 2. A t-test showed that this difference was significant.
- Outside of GSR, there is no statistically significant difference in recorded emotions between Game 1 and Game 2.

While our quantitative analysis found that GSR data correlated with the SAM data, we found this further inconsistencies with our qualitative findings. Though participants self-reported they were happier, less excited, and more in control in Game 2, this feedback was not always consistent either the SAM or the iMotions data.

Some of our Qualitative findings are highlighted below:

- Participants self-reported they were happier, less excited, and more in control in Game 2.
- When P2 lost the game in Junior year in game 1, she self-reported being unhappy and less excited, the biometric data showed spikes in contempt, disgust, and surprise. P2 also lost the game in senior year during the second attempt, she self-reported being unhappy, less excited, and have less control. P2's verbal feedback was thus **consistent** with the other measures.
- During the graduation period of game 1, P6 self-reported being unhappy on SAM Valence while biometric data implied that about 50% of the game time, she was expressing the emotion of Joy. Contradictory, in the next segment, the pre-college period of game 2, she reported much higher happiness comparing to previous segment while biometric data captured much lower emotion of Joy. P6's verbal feedback was thus **inconsistent** with the other measures.

The inconsistency in these findings highlight why it is so important to triangulate multiple sources of participant emotion data to get a fuller picture of participant emotions during qualitative studies. Though the study was too small to develop which system produced "truest" results, in concert they can be used together to highlight areas of interest for further exploration.

Quantitative Analysis - Comparing SAM, GSR, and iMotions

We identified two statistically significant findings when plotting the participant's self-reported SAM against iMotions and GSR data in order to take a closer look at how each reported user emotions.

1. **Arousal x GSR:** Participants showed more arousal in higher peaks/min as measured by GSR in Game 2. A t-test showed that there was a significant difference between Game 1 and Game 2. During Game 1, there was a positive correlation between GSR and SAM Arousal. As participants GSR increased, participants reported more arousal (excitement). (See figure 2)

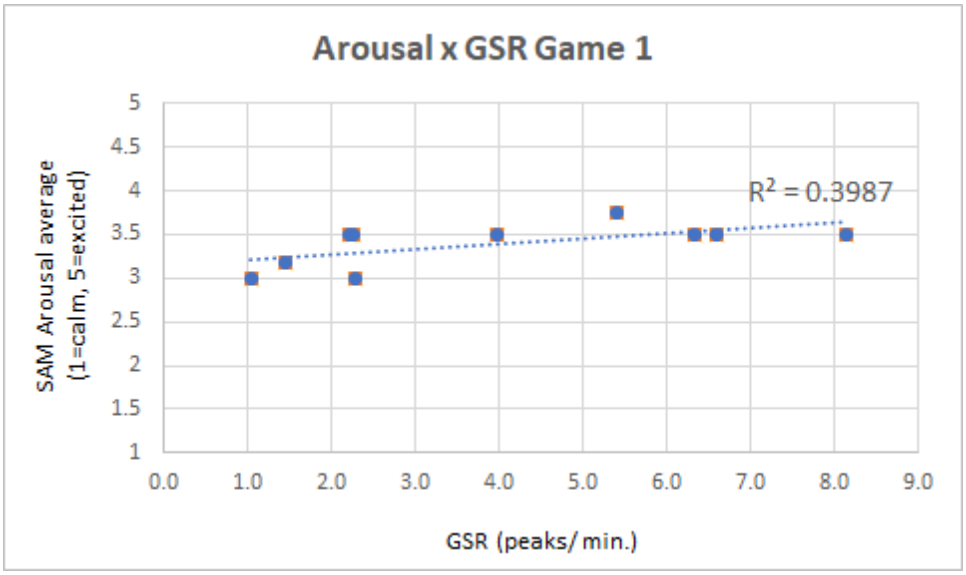

Figure 2 *Arousal x GSR Game 1*

During game two, there was a negative correlation between GSR and SAM Arousal. As participants GSR increased, participants reported less arousal (excitement). (See figure 3).

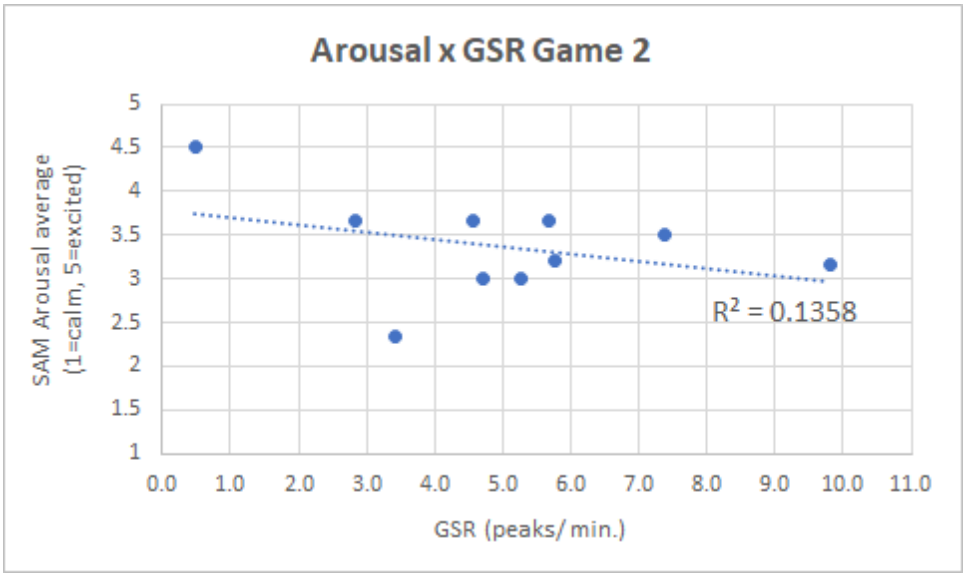

Figure 3 *Arousal x GSR Game 2*

2. **Valence x GSR** Participants reported a difference in valence between Game 1 and Game 2. A t-test showed that this difference was significant. During game 1, as iMotions picked up more joy, participants self-reported happiness decreased. During game 2, as iMotions picked up more joy, participants self-reported happiness remained flat. (See figure 4 & 5).

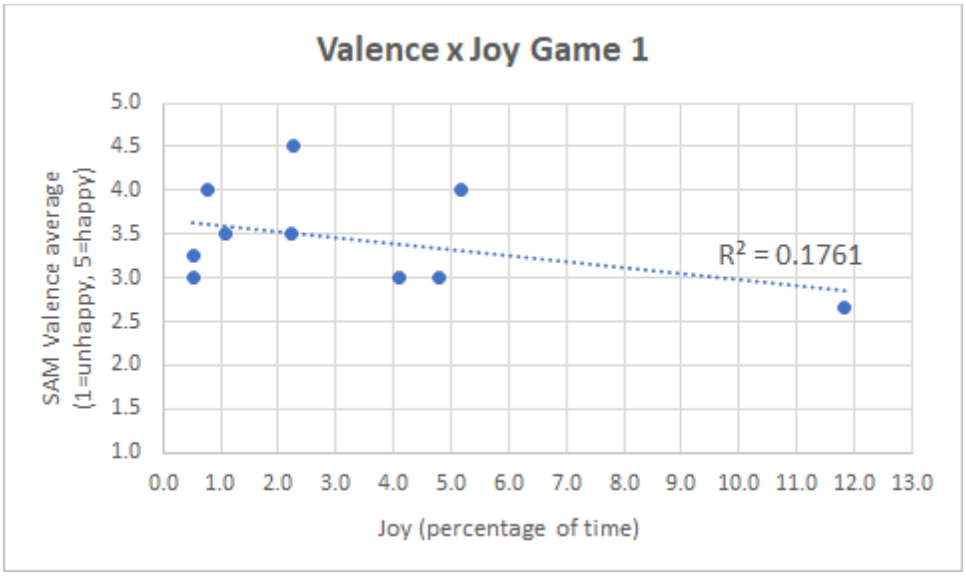

Figure 4 *Valence x Joy data from iMotions in Game 1*

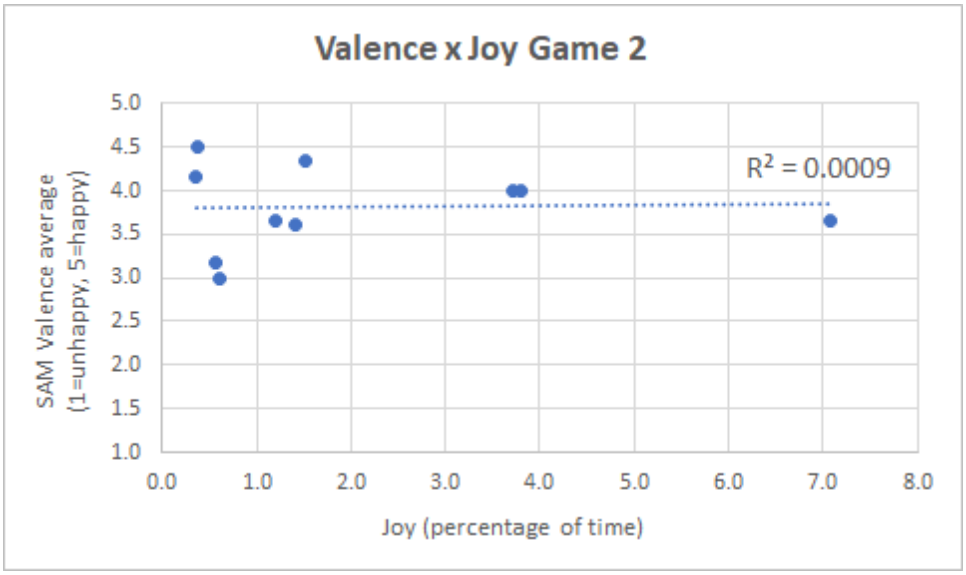

Figure 5 *Valence x Joy data from iMotions in Game 2*

Other Trends

Outside of GSR, there is no statistically significant difference in recorded emotions between Game 1 and Game 2. We did however notice several other trends.

Dominance x GSR. A t-test did not show statistical significance of difference between Dominance and GSR in Game 1 vs, game 2, however both were positively correlated.

In other words, as Arousal increased, participants self-reported an increase in feeling in control in both game 1 and game 2 (see figure 5 and 6).

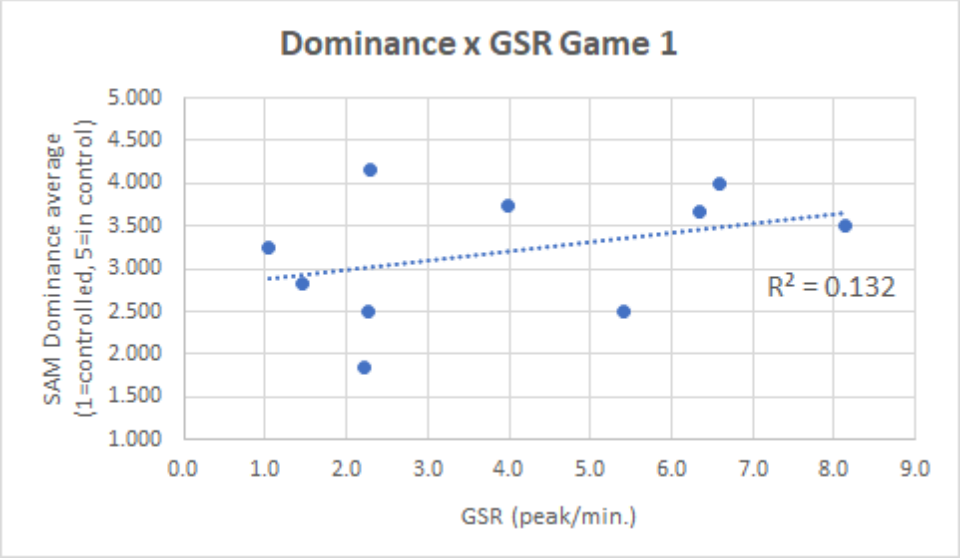

Figure 5 *Dominance x GSR in Game 1*

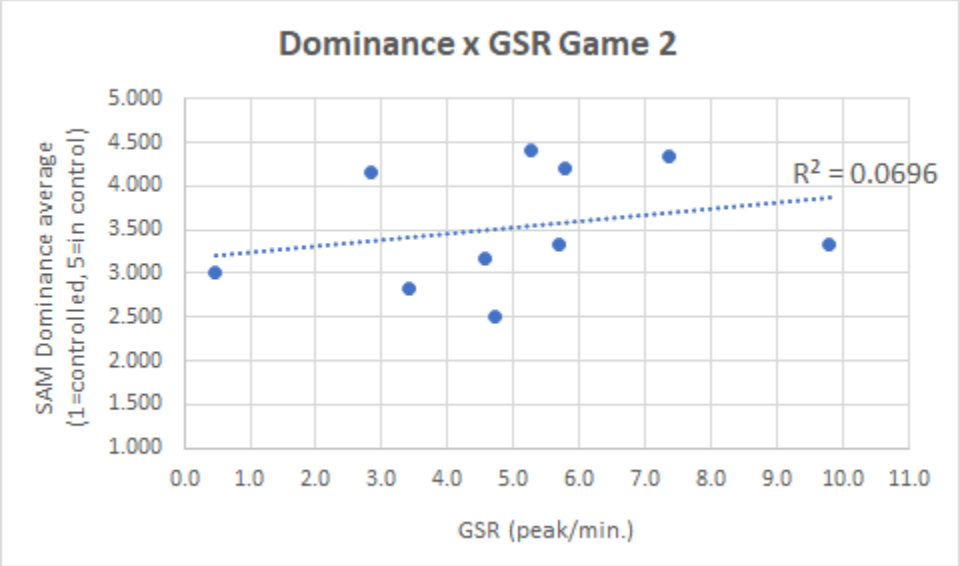

Figure 6 *Dominance x GSR in Game 2*

Additional analysis revealed several moderate (.6 <p <.7) and strongly (.7 < p) correlated data points between the SAM, Emotions, and GSR as outlined below.

Strongly correlated data:

During the Freshman Year first attempt, valence was strongly correlated with both Disgust (.762) and Sadness (.764). This is interesting to note and further supports the idea that emotions are difficult to assess.

During the Sophomore Year first attempt, valence was strongly correlated with anger (.742). The amount of anger detected positively correlated with people's self-reported happiness.

During the Sophomore year first attempt, valence was strongly correlated with GSR. When GSR detected an increase in arousal it was strongly correlated an increase in participants self-reported sense of confidence or being more in control (average 3.2).

During Junior year first attempt, arousal was strongly correlated with Surprise (.733). As increased surprise was detected, increased arousal was reported.

At the Graduation screen during the second attempt when participants are reviewing their final numbers, arousal was strongly negatively correlated with surprise (-.717). As surprise as detected in iMotions decreased, self-reported excitement increased.

Order	Stimuli Name	SAM vs Emotion/GSR	(+/-)	Value
2	Freshman	Valence vs Sadness	Positive	+0.764
3	Sophomore	Dominance vs GSR	Positive	+0.763
2	Freshman	Valence vs Disgust	Positive	+0.762
3	Sophomore	Valence vs Anger	Positive	+0.742
4	Junior	Arousal vs Surprise	Positive	+0.733
10	Graduation 2	Arousal vs Surprise	Negative	-0.717

Key
SAM Dominance
SAM Valence
SAM Arousal

Moderately correlated data:

Throughout the test the results showed a positive relationship between the SAM Dominance and GSR. While being only moderately correlated, these positive correlations demonstrate that the SAM and iMotions data corresponded consistently, which is of note.

We then completed a T-test analysis of all variables of the two games. Of those variables only GSR and SAM Valence averages were we able to measure a difference. This provides evidence that the data collected was not simply by chance but provided a measurable difference.

Order	Stimuli Name	SAM vs Emotion/GSR	(+/-)	Value
8	Junior 2	Valence vs Fear	Positive	+0.675
5	Precollege 2	Arousal vs Anger	Positive	+0.667
5	Precollege 2	Arousal vs Sadness	Positive	+0.666
3	Sophomore	Valence vs Sadness	Positive	+0.645
8	Junior 2	Dominance vs Disgust	Positive	+0.607
8	Junior 2	Dominance vs Sadness	Positive	+0.607
6	Freshman 2	Dominance vs GSR	Positive	+0.606
2	Freshman	Dominance vs Sadness	Negative	-0.610
2	Freshman	Valence vs GSR	Negative	-0.611
2	Freshman	Dominance vs Disgust	Negative	-0.620
6	Freshman 2	Arousal vs Fear	Negative	-0.629
6	Freshman 2	Valence vs Surprise	Negative	-0.643
7	Sophomore 2	Arousal vs Contempt	Negative	-0.643
6	Freshman 2	Valence vs Joy	Negative	-0.680

Key
SAM Dominance
SAM Valence
SAM Arousal

Qualitative Analysis - Verbal Feedback vs. SAM Measurement

In addition to the SAM, Participants gave verbal feedback following each section of the gameplay. Overall, during game 2, participants self-reported they were happier, less excited, and more in control in Game 2. We have provided additional sampling of quotes below along with the corresponding SAM and biometric data for each section to give a better understanding of participant experience.

These findings indicate that a participant's words did not always match up with what they self-reported nor what emotions iMotions highlights. We have noted where these quotes are consistent or inconsistent with the SAM.

One important note in reviewing these examples is that we inverted the original SAM A to line up with the GSR (the higher the number, the higher the reported happiness and excitement).

Examples of Verbal Feedback Consistent with SAM:

P2: When P2 lost the game in Junior year in game 1, she self-reported being unhappy and less excited, the biometric data showed spikes in contempt, disgust, and surprise. P2 also lost the game in senior year during the second attempt, she self-reported being unhappy, less excited, and have less control.

"I feel good; it is encouraging when it says that I'll save money"

- P2, Sophomore Year 1(4 - somewhat happy, consistent)

- Joy (12.781), Surprise (1.81), contempt (0.235) disgust, (.29) and fear (.217) were apparent.

P5: When P5 lost the game during the Junior year in game 1, he self-reported being unhappy, less excited and has less control, and the biometric data showed huge spike in contempt.

"Not as much in control. I didn't expect not studying abroad would make me depressed"

- P5, Junior Year 1 (2 - somewhat unhappy, consistent)

Contempt (28.233) and Surprise (.21) were apparent

P9: lost the game during the Junior year in game 1, she self-reported with higher excitement and biometric data confirmed that with higher reading of Joy.

P10: lost the game during the Junior year in game 1, she self-reported being unhappy, and biometric data confirmed that by showing spikes in anger, disgust, sadness and surprise.

Examples of Verbal Feedback not consistent with SAM:

P6: During the graduation period of game 1, P6 self-reported being unhappy on SAM Valence while biometric data implied that about 50% of the game time, she was expressing the emotion of Joy. Contradictory, in the next segment, the pre-college period of game 2, she reported much higher happiness comparing to previous segment while biometric data captured much lower emotion of Joy.

Game 1:

"I feel bad; my debt was higher than my salary but income is growing."

- P6, Graduation 1 (SAM Valence: 2 – somewhat unhappy)

Joy was at (50.164), the highest of all joy numbers, fear was at .437

Game 2:

"Kind of unhappy, I can see my debt going up which isn't fun."

- P6, Senior Year 2, (SAM Valence: 3 - neutral, was somewhat unhappy (2) in game 1, **inconsistent**)

- Fear (2.298) is the only registering emotion vs. Fear (8.883) and Joy (5.585) registering in game two.

P1: "A little nervous, trying not to mess up; I'm a little excited"

- Freshman Year 1 (SAM Arousal: 3 - neutral, **inconsistent**)

No emotion registered.

GSR was at 1.17

P3: "I feel moderately in control. The grad school decision was good for higher earnings."

- p3, Graduation 2 (2 - somewhat controlled, was in control (5) in game 1, **inconsistent**)

- Joy (13) and contempt (8.43) were the only emotions registered

Example of Verbal Feedback not consistent with SAM or iMotions

P10: "Sad I can't study abroad because it costs so much, but it happens"

- P10, Junior Year 2 (SAM Valiance: 4 - somewhat happy, i**nconsistent**)

- Fear (1.601) and Anger (.123) were the only emotions registering.

Limitations

There are several notable limitations of the study which impacted the findings. All of which limitations would have to be addressed in a larger study:

1. **Study Size.** There are only 10 participants in this test. True statistically significant summative tests require 25 or more participants. The statistics we have gathered, while descriptive, cannot be taken as statistically significant across a greater pool of users. We can only draw conclusions about behavior of the group we have studied.
2. **Variability of gameplay.** No two player games were exactly alike. The game was designed to present different outcomes for the same decision, such as how much financial aid was awarded in a given year. Also some sections have a higher demand for human motor ability, which affects result based on player success.
3. **Completion of the game is not guaranteed.** Not everyone was able to finish the whole game. Some participants failed to balance one of the three objects in the game (focus, connection, and happiness) at different points of the game (Sophomore year, Junior year, etc.)
4. **Ambiguity of facial expressions.** iMotions reads facial expressions to detect emotion. However, emotions are known to be expressed differently across individual situations as well as cultures. Happiness is particularly ambiguous – though smiling typically indicates happiness it can also be a signal of acquiescence or is merely a biological reaction to being in a crowd (Barrett et al., 2019, p. 20-21).

Conclusion

The results of our research conclude that using both in concert with one another can help triangulate areas of pain and get a more complete understanding of participant emotions. GSR shows a strong correlation with results and supplementing it with SAM makes it even stronger. The potential for emotion detection is significant given the new technologies. Important to keep in mind is that there was correlation between GSR and SAM, and that these correlations often trended with iMotions findings, there is not enough statistical significance or evidence to support which one best tracks participant emotions. Next steps would include a validation study with at least 25 participants in a very controlled study to work towards greater understanding of the topic.

In the meantime, we recommend using iMotions as a third method of triangulation with GSR and Self-Report to dig deeper into participant emotional reactions to studies. The real-time emotion reading which iMotions delivers during a qualitative study that has been designed to incorporate both emotion tracking and think-aloud provides great flexibility for researchers. With at least one researcher tracking spikes in iMotions analysis, they can prompt the moderator to dig deeper into self-reported data if the initial SAM ratings seem inconsistent with the recent playback. This “Reflective Think Aloud” could help researchers pinpoint the exact spot someone had the reaction and follow up with further details. We look forward to incorporating this methodology, and other best practices from these findings, in future client studies

References

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest,* 20(1), 1–68. https://doi.org/10.1177/1529100619832930