Exploring the Nature of Affect, Confusion and EDA in Learning Game Play

Jen Groff
Trung Tran
MAS.630
Dec 2013

Introduction

There is long-held connection between emotion and cognition. This is known and discussed in the learning sciences and in education, but because it is difficult to observe, measure, direct, etc., it often gets glossed over and compartmentalized as the caveat, “yes you need to motivate and engage the learner” but then all subsequent attention is focused on the learning content.

‘Basic’ versus ‘Deep’ Learning

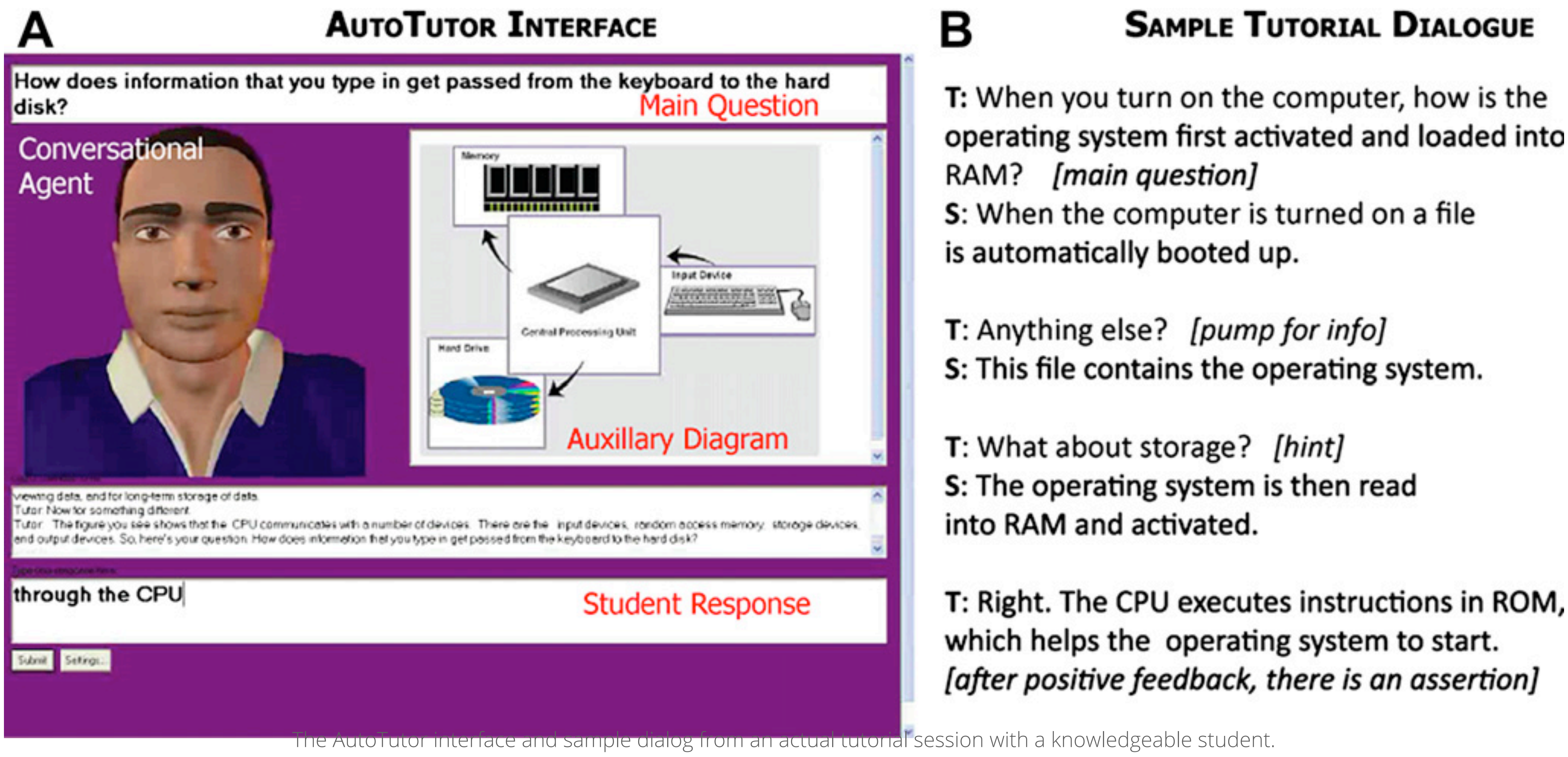
It is only more recently that we have begun to explore a more nuanced analysis of emotions and learning, particularly as it relates to deep learning, or complex learning—where the concept to be learned is not straightforward such as learning multiplication tables, but a messy conceptual space that requires the learner to push and pull challenging material over time. The goal is to explore not only what emotional states occur in these processes but how they change or interplay over the course of a learning experience. Sidney D’Mello and Art Graesser have been leaders in much of this work, through their explorations in intelligent tutoring systems and simulations. We decided to use their work as a starting point for our project in the context of educational games.

Background

Background: Games & Learning

Though the field of educational gaming is more than 30 years old, it's really only in the last decade that it has begun to develop real rigor. It's an emerging and booming field, with intense interest from multiple areas on how to best leverage games as a learning intervention and even as powerful assessments. This is in large part due to the many unique traits to game--including their feedback structures, scaffolded experiences, visuals and playful mechanics--all which contribute to high rates of engagement and motivation.

However many questions still remain about to best design games to leverage these qualities. Due to their dynamic and often open-ended nature, it can be difficult to study the learning and other cognitive activities at a micro-level across the game. Learning game designers often discuss the notion of ZPD (Zone of Proximal Development) and Flow as key elements that games leverage, however truly understand how to leverage these and how they actually play out in games is very difficult.

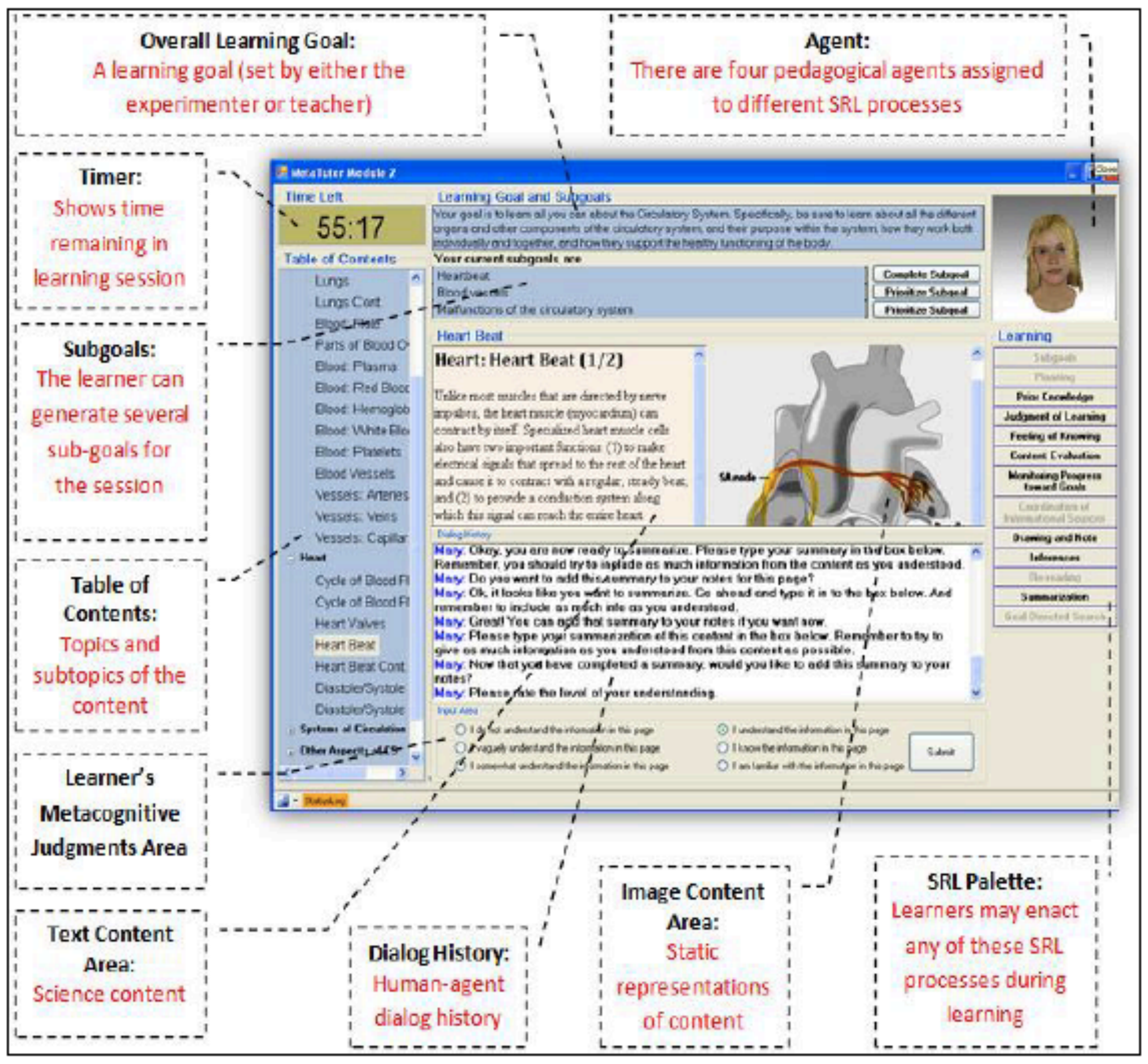


The AutoTutor interface and sample dialog from an actual tutorial session with a knowledgeable student.
source: D'Mello & Graesser, 2012

Conceptual Change and Confusion
A prominent element in the learning sciences is Conceptual Change theory (Strike & Posner, 1992), which describes how individuals are continually assimilating new information into existing knowledge structures (e.g., existing schemas or mental models) when they are engaged in a complex learning task.

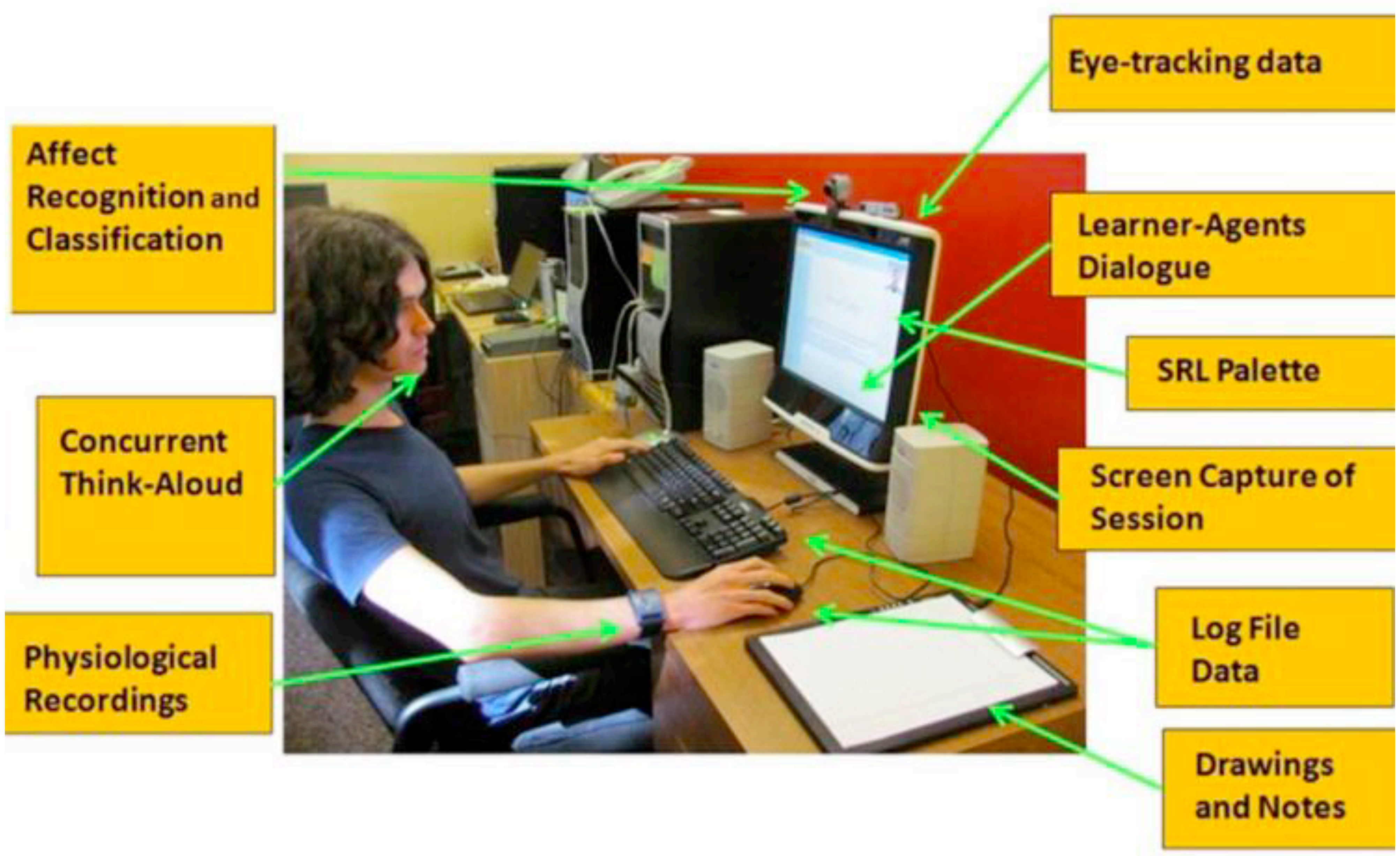
When new information that is encountered is discrepant or in conflict with prior knowledge, the autonomic nervous system increases in arousal, and the individual experiences a variety of possible emotions as they try to reconcile this discrepancy (Stein & Levine, 1991).

Affect & Intelligent Tutoring Systems
Using intelligent tutors as a context for exploring confusion and affect, D'Mello et al. (2014) found that when interacting with the system, learning gains were positively correlated with confusion and engagement/flow, negatively correlated with boredom, and that confusion was the only emotion that significantly predicted learning. They explain, "we know that activities of the sympathetic nervous system increase when there is cognitive disequilibrium compared to a neutral state" (Graesser & D'Mello, 2011, p. 13).

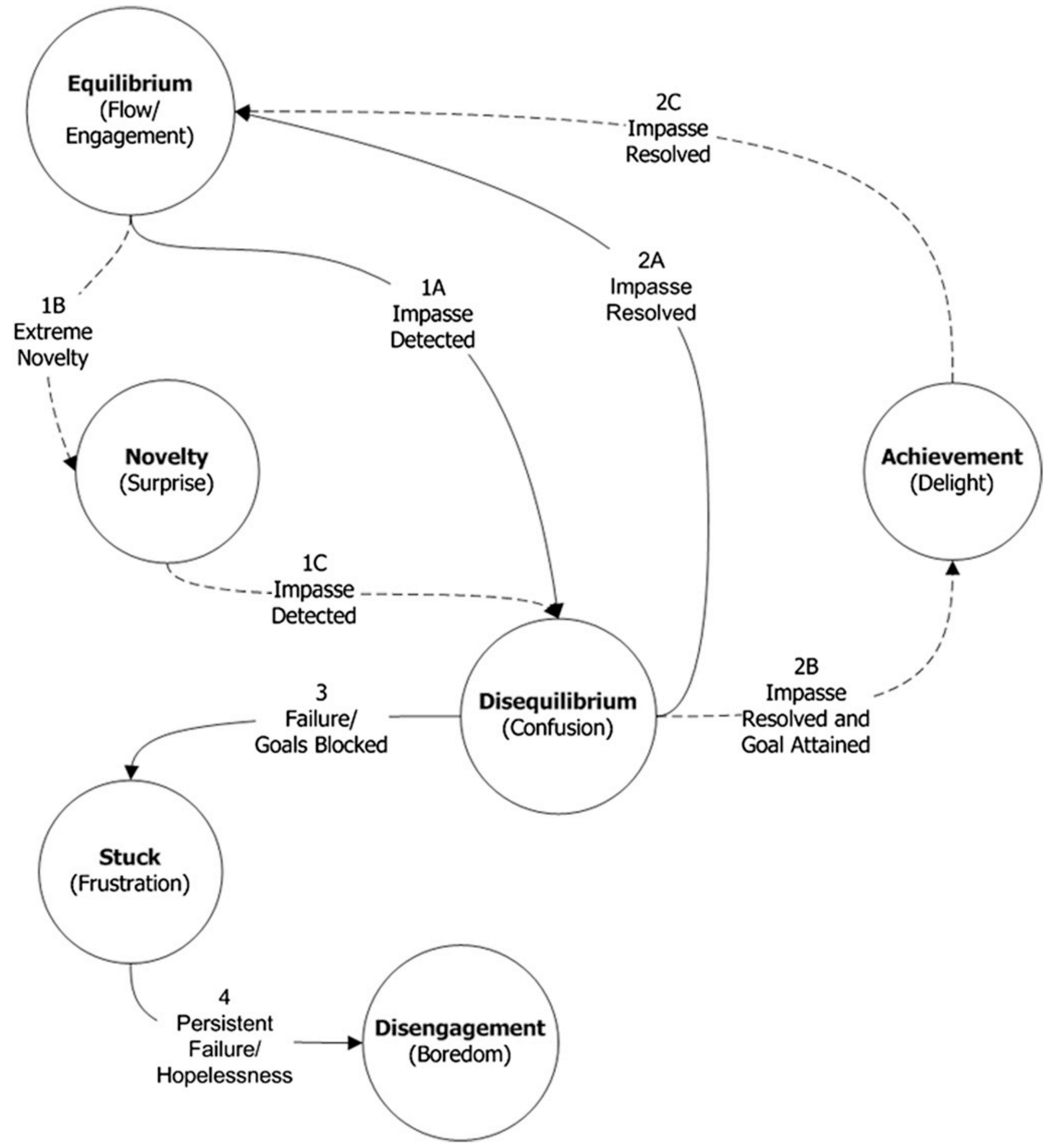


Annotated screenshot of MetaTutor
source: Azevedo et al., n.d.

Roger Azevedo's group at McGill University in Quebec has developed MetaTutor, an intelligent multi-agent tutoring system designed to detect, track, model, and foster self-regulated learning. MetaTutor is both an instruction and research tool. As a research tool, "MetaTutor is capable of measuring the deployment of self-regulatory processes through the collection of rich, multi-stream data including: self-report measures of SRL, on-line measures of cognitive and metacognitive processes (through the use of concurrent think-alouds), dialogue of agent-student interactions, natural language processing of help-seeking behavior, physiological measures of motivation and emotions, emerging patterns of effective problem solving behaviors and strategies, facial data on both basic (e.g., anger) and learning-centered emotions (e.g., boredom), and eye-tracking data regarding the selection, organization, and integration of multiple representations of information (e.g., text, diagrams)" (Azevedo et al., n.d.).



Experimental set-up for the collection of cognitive, metacognitive, and affective processes during learning with MetaTutor.
source: Azevedo et al., n.d.



D'Mello and Graesser's hypothesized model of affect dynamics.
source: D'Mello & Graesser, 2012

A Potential Model for Affective Stages in Deep Learning
D'Mello and Graesser have proposed a possible model to explain the affective states that emerge during deep learning activities. They argue that,

> *"The model predicts that learners in a state of engagement/flow will experience cognitive disequilibrium and confusion when they face contradictions, incongruities, anomalies, obstacles to goals, and other impasses. Learners revert into the engaged/flow state if equilibrium is restored through thought, reflection, and problem solving. However, failure to restore equilibrium as well as obstacles that block goals trigger frustration, which, if unresolved, will eventually lead to boredom."*

~ D'Mello & Graesser, 2012

D'Mello and Graesser have sought to demonstrate support for their model through two small studies, in which affective states were tracked during a 30 minute session with the tutor, and their analysis demonstrated the presence of confusion—engagement/flow, boredom—frustration, and confusion—frustration oscillations.

Confusion, Flow and Educational Games
Graesser and D'Mello go further to make the connection between their model and flow: "Our best interpretation of Csikszentmihalyi's flow state is that it is an emergent affect state from a set of smaller-scale cycles that involve modest challenges, high engagement, timely achievement, and delight" (2011, p. 14). This model aligns well with dynamics often observed in game play with educational games. However, we believe that the heart of this model is very nuanced, and the way the dynamics of it play out will vary tremendously depending on the context, content, learner's goals, learner's desired modalities, and many other factors—and it is the very nature of these dynamics that are most critical to explore and seek to understand.

Motivations & Questions

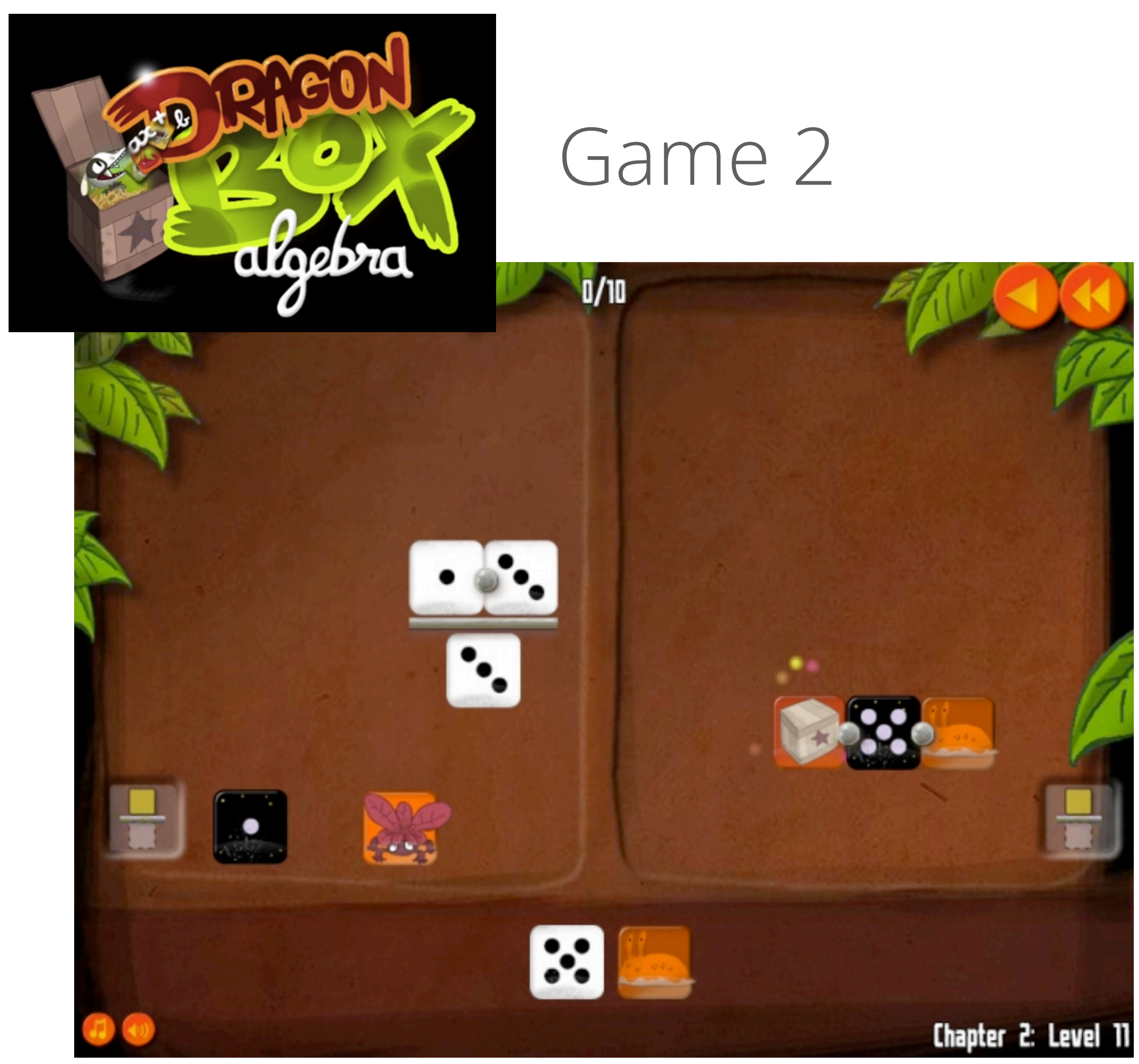
An Exploratory Study of Affect in Educational Game Play
Given the very different nature of dynamics and interplay with educational games as compared to cognitive tutors – as well as the vast differences with educational game genres themselves – we chose to conduct an exploratory study on the nature of arousal and confusion in educational game play.

- Can we observe confusion/resolve cycles in game play?
- Can we distinguish between cognitive confusion and game play confusion?
- What can we observe in relation to player self-reports?

Experiment Design

Game 1

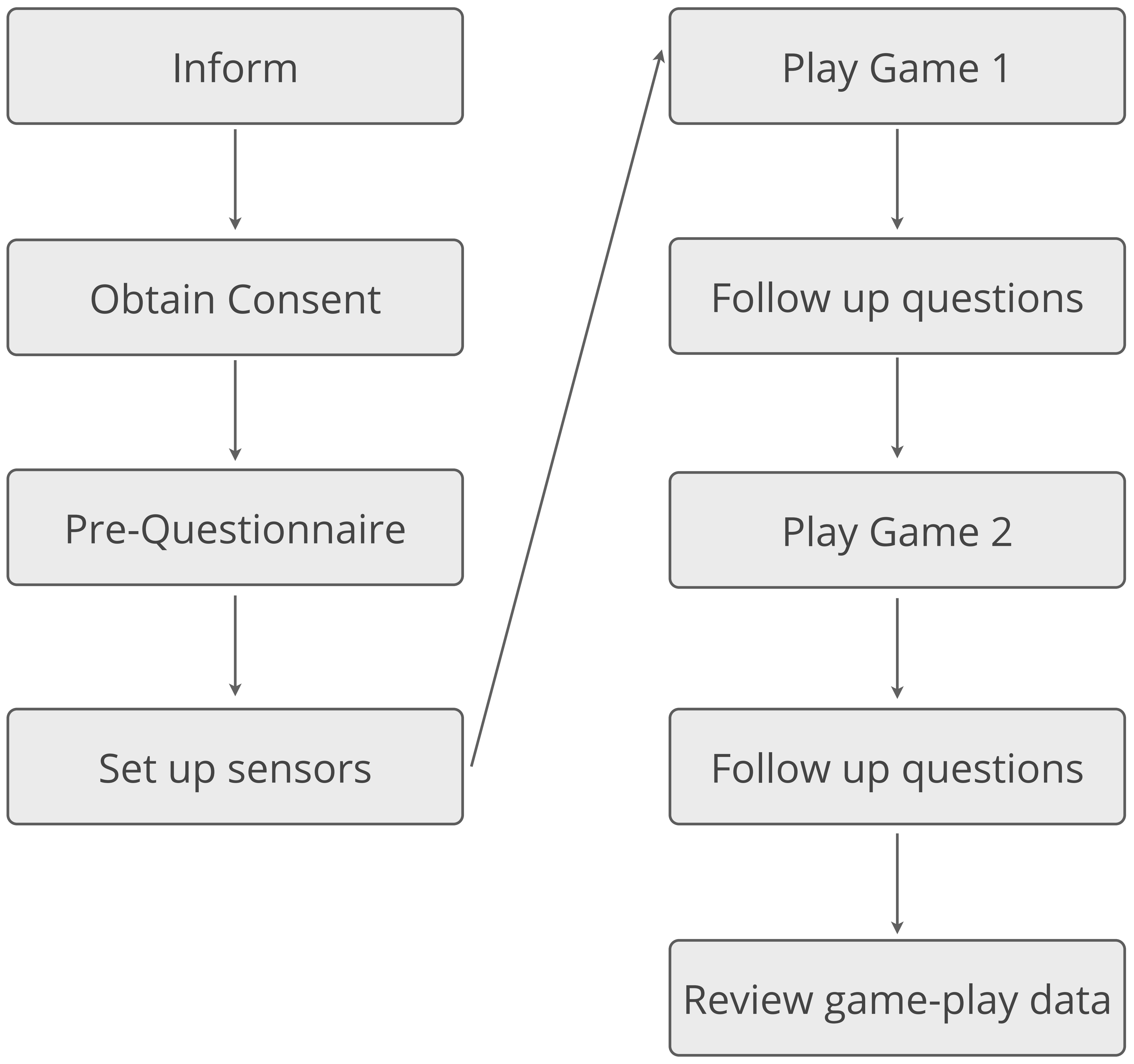
We chose two educational games, for deliberate reasons. First, The Radix Endeavor developed by the Educational Arcade group at MIT, is a MMO (Massively Multiplayer Online) game that explore STEM concepts. It is a 2D, open-ended world that learners can explore and take on different quests with unique learning goals.



Game 2

We chose the Algebra quests within the game to be the focus for our study. Therefore we wanted to use an additional learning game with similar learning goals but of a much different genre. DragonBox is highly popular game that scaffolds algebraic thinking in players. A casual puzzle game, the dynamic of play is much different than Radix.

Protocol



We chose to collect both quantitative and qualitative data from game-playing experience. Initially, all participants received necessary information and instruction on their role in the experiment and were asked to provide written consent. A pre-questionnaire is conducted to gather basic demographics and game experiences. Participants then proceed to play Radix with the objective of carrying out a quest to learn algebra. Upon finishing their Radix session, participants are asked for evaluative opinion about their experience. Subsequently, a session with Dragon Box follows the same structure. In the final stage of the experiment, researchers sit down and review gameplay footages with participants in order to obtain qualitative interpretations of specific in-game EDA stimulants.

Protocol: questions

Setup

Age
Gender
I play video games:

A	B	C	D
Never	1-2 hours per week	3-6 hours per week	7+ hours per week

What types of games do you generally play?

I'm excited by learning via video games:
1--------------------------------5
Not at all Absolutely!

You are about to play 2 different games about algebra concepts.
Rate yourself on a scale from 1-5 on your ability in algebra:
1--------------------------------5
Terrible Expert

Before you begin, just based on initial looks, how excited are you to play this game?
1--------------------------------5
Not at all Can NOT wait!

Before you begin, just based on initial looks, how excited are you to play this game?
1--------------------------------5
Not at all Can NOT wait!

Play Radix

What did you like about the game?
What didn't you like?

Before you begin, just based on initial looks, how excited are you to play this game?
1--------------------------------5
Not at all Can NOT wait!

Play DragonBox

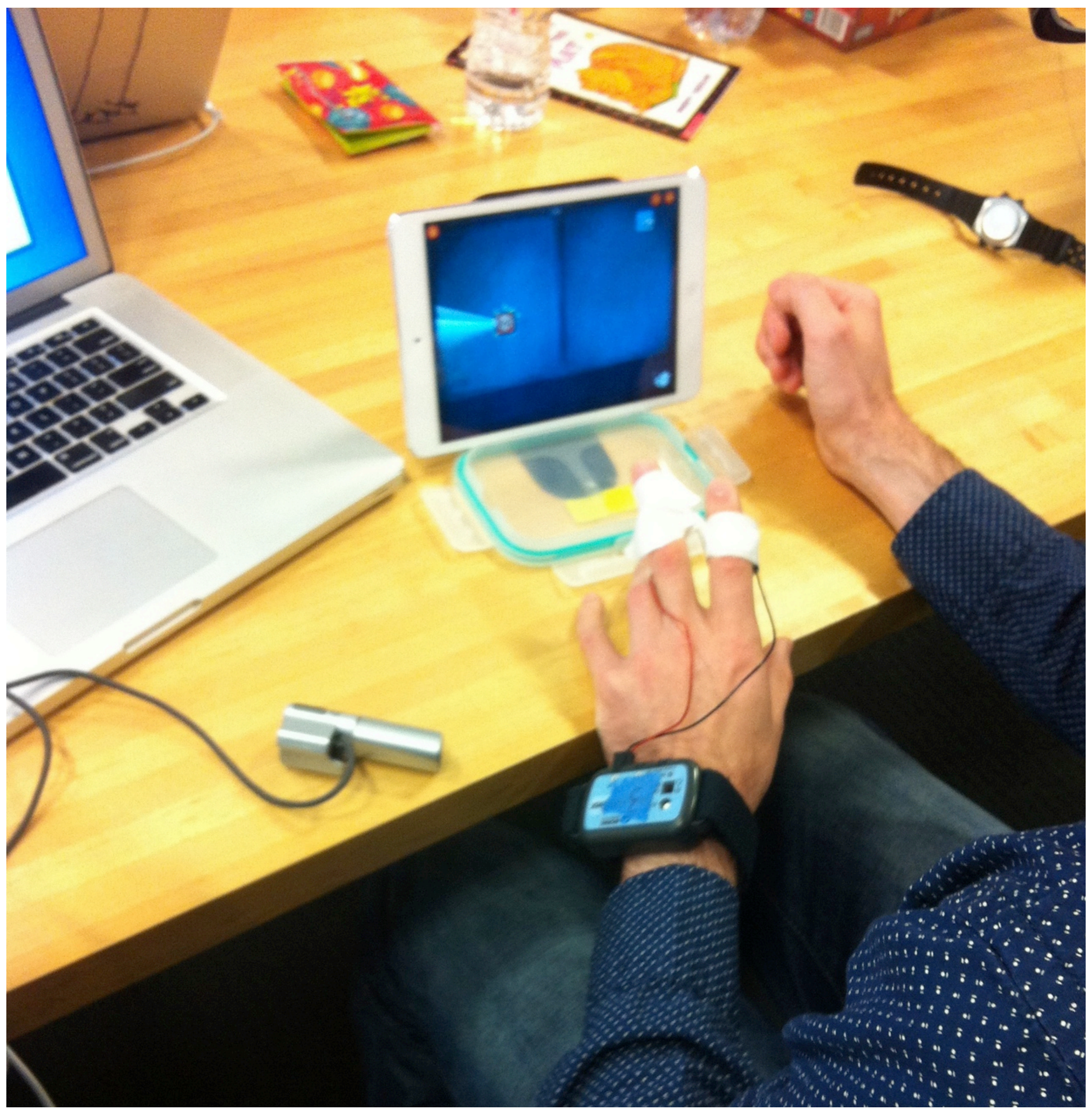
What did you like about the game?
What didn't you like?

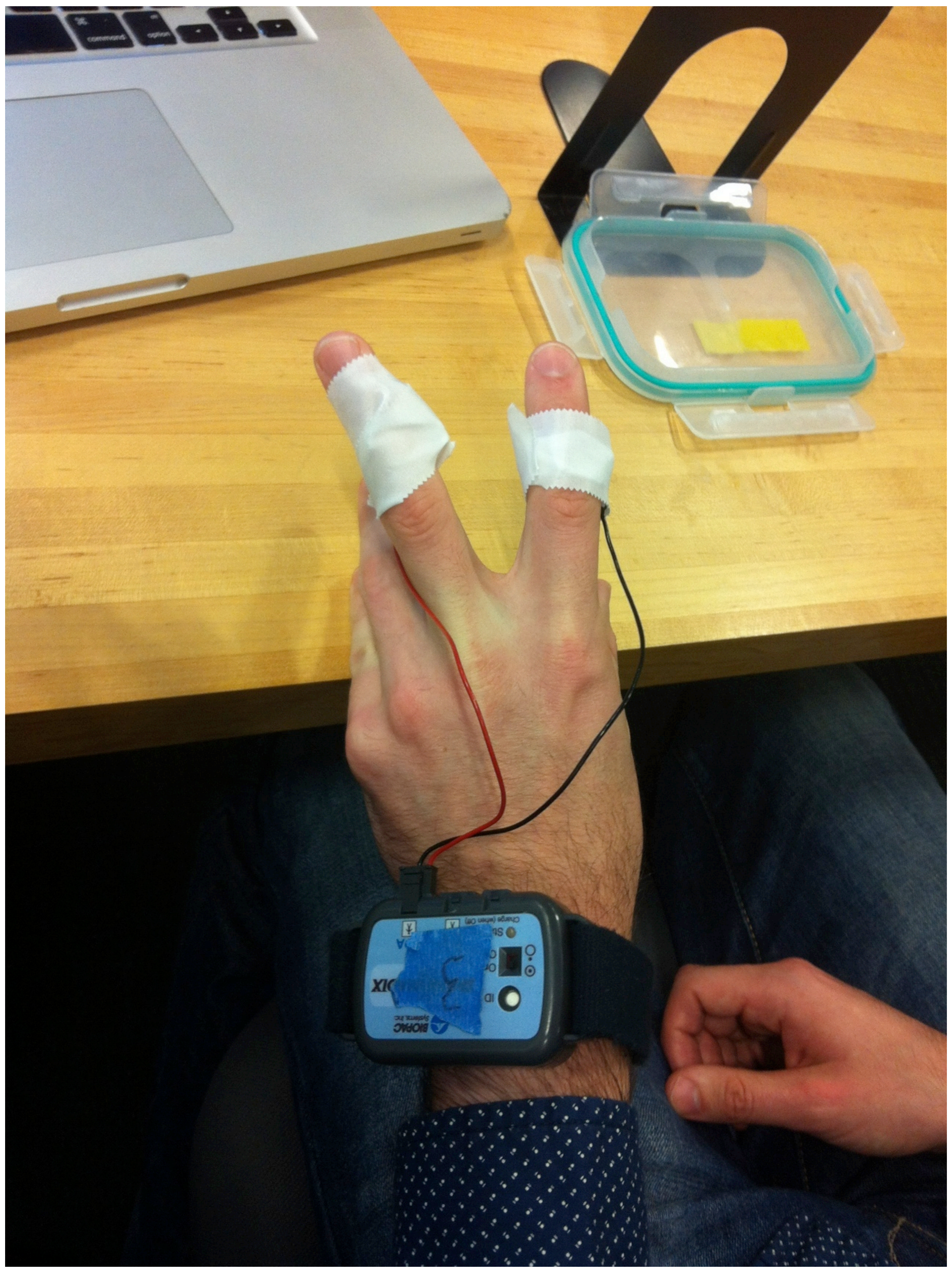
Wrap-Up

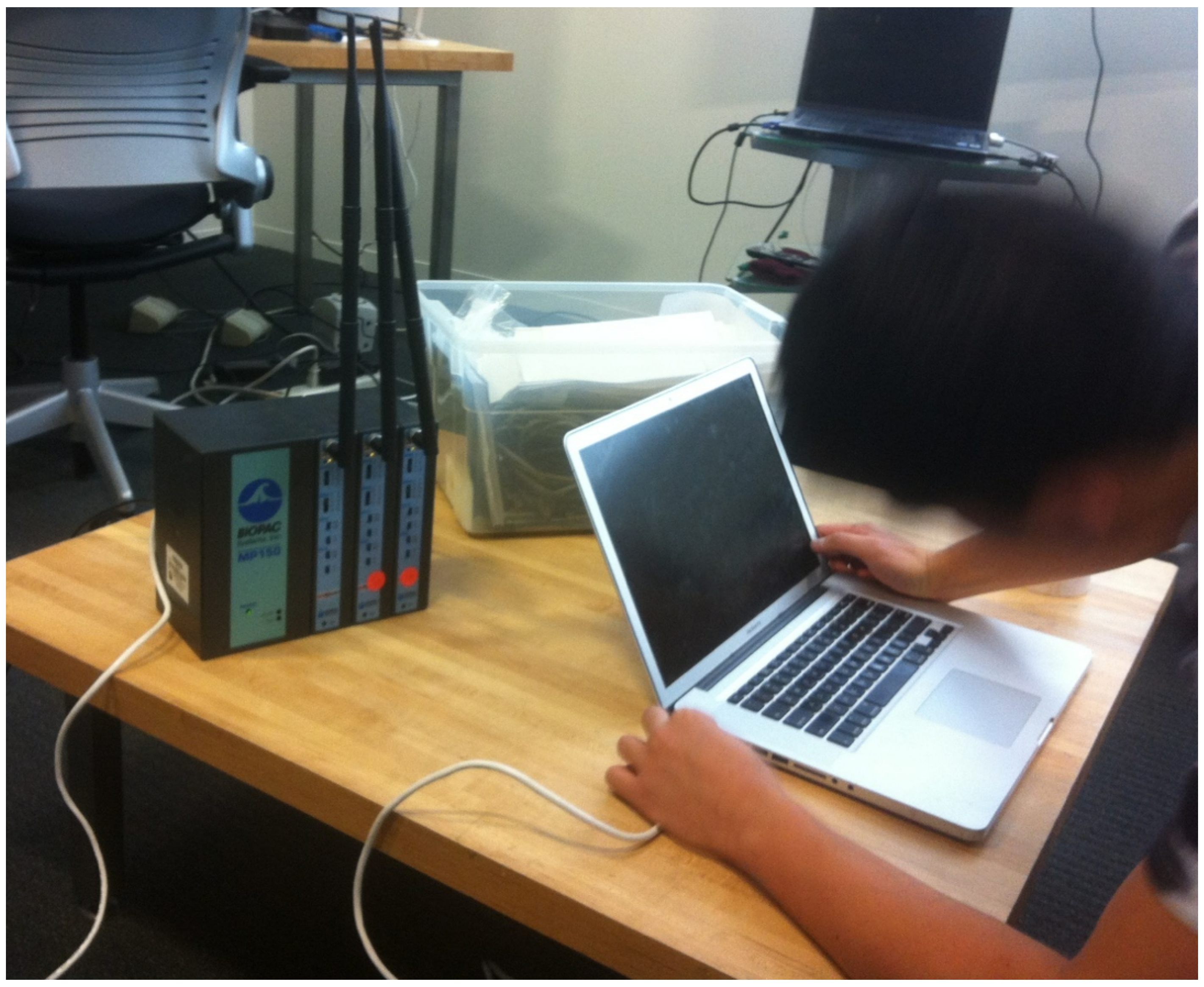
Which game did you like better? Why?
What would have made you played Radix longer?
What would have made you played DragonBox longer?

Besides EDA measurement and gameplay review self-report, the pre-questionnaire is an important data collecting method. First, general demographic such as age, gender, and familiarity with video game are collected. Additionally, background information on participants' opinion about game learning, subject-matter, and initial engagement are collected. After each game play session, participants will reflect on their in-game experiences through specific likes and dislikes. Finally, to conclude the experiment, wrap-up questions provided participants with an opportunity to provide reflective feedback of both games on a comparative basis. While such questions do not directly address our research questions, they can potentially be insightful in understanding unique patterns in EDA signals.

Affective Technologies







Since we needed to align video recordings of the participant game play with the EDA data, we chose to use the BIOPAC AcqKnowledge data acquisition analysis sensor system and software, owned by Elliott Hedman in the lab, who has used this system extensively to explore affect across a number of products and technologies. The system include a sensor worn on the wrist and attaches to middle pads of the index and middle fingers, as well as receiver that is able to collect the data from the sensor while displaying it simultaneously on the AcqKnowledge software on the computer.

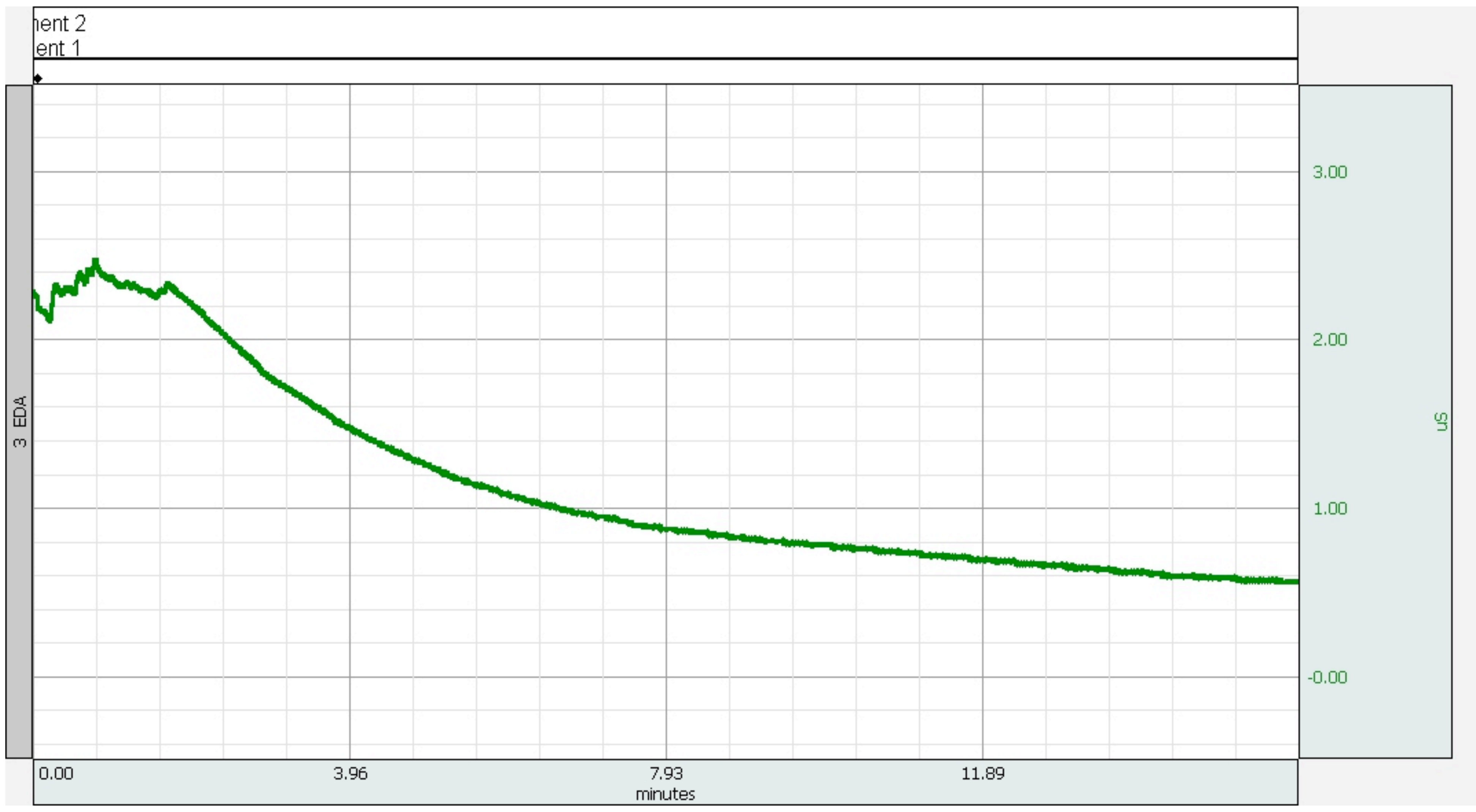
Participants

Participant 1	Participant 3
28 male 3-6 hours per week casual games	29 female 1-2 hours per week mobile, puzzle games

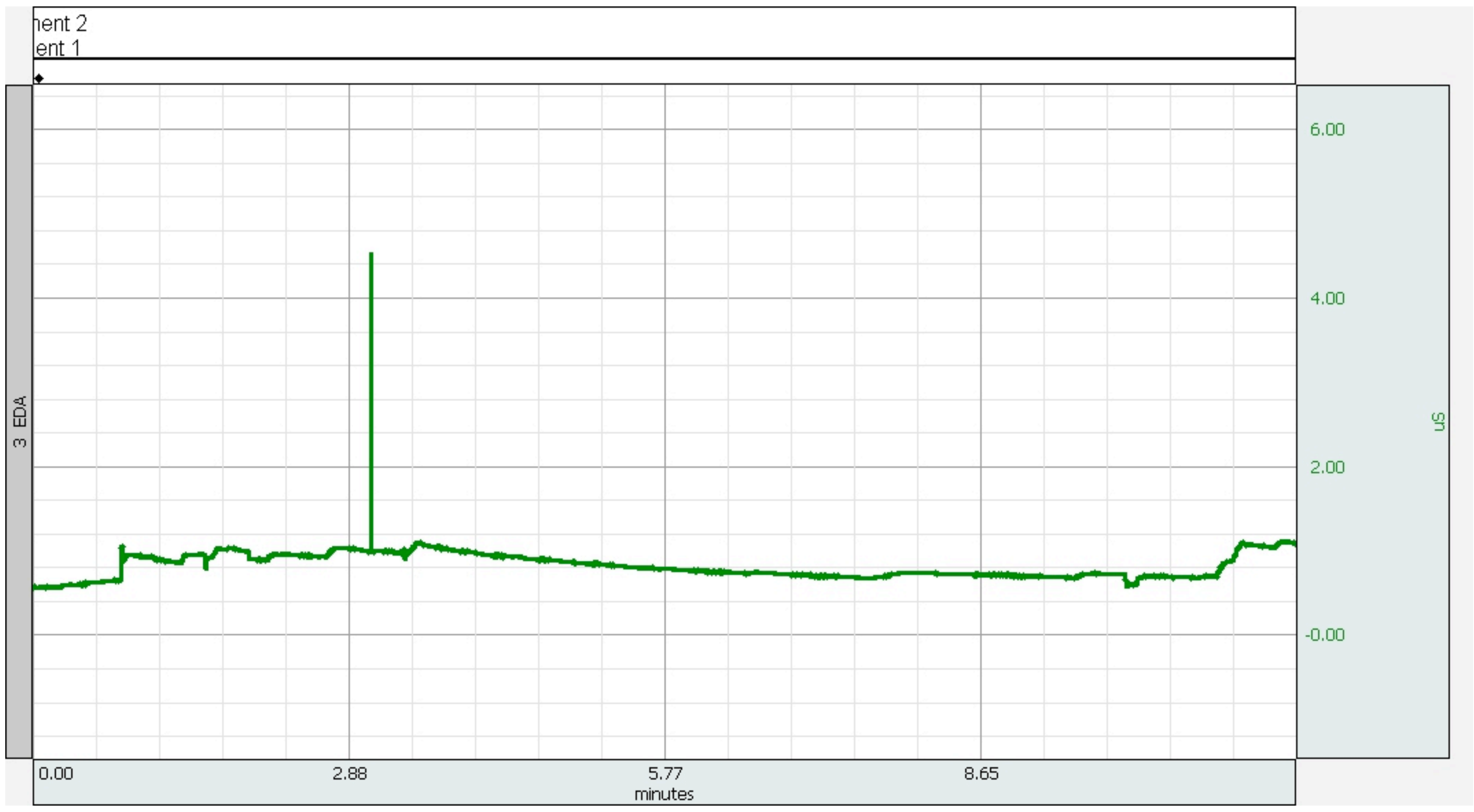
Participant 2	Participant 4
27 male 3-6 hours per week strategy, FPS games	32 male 3-6 hours per week sandbox, puzzle

Participants were recruited from the MIT community. While participants fall into a similar age range and have similar game-play habitudes, they do have different tastes in video games and this preference will be considered as we review the game play data.

Findings: Participant 1



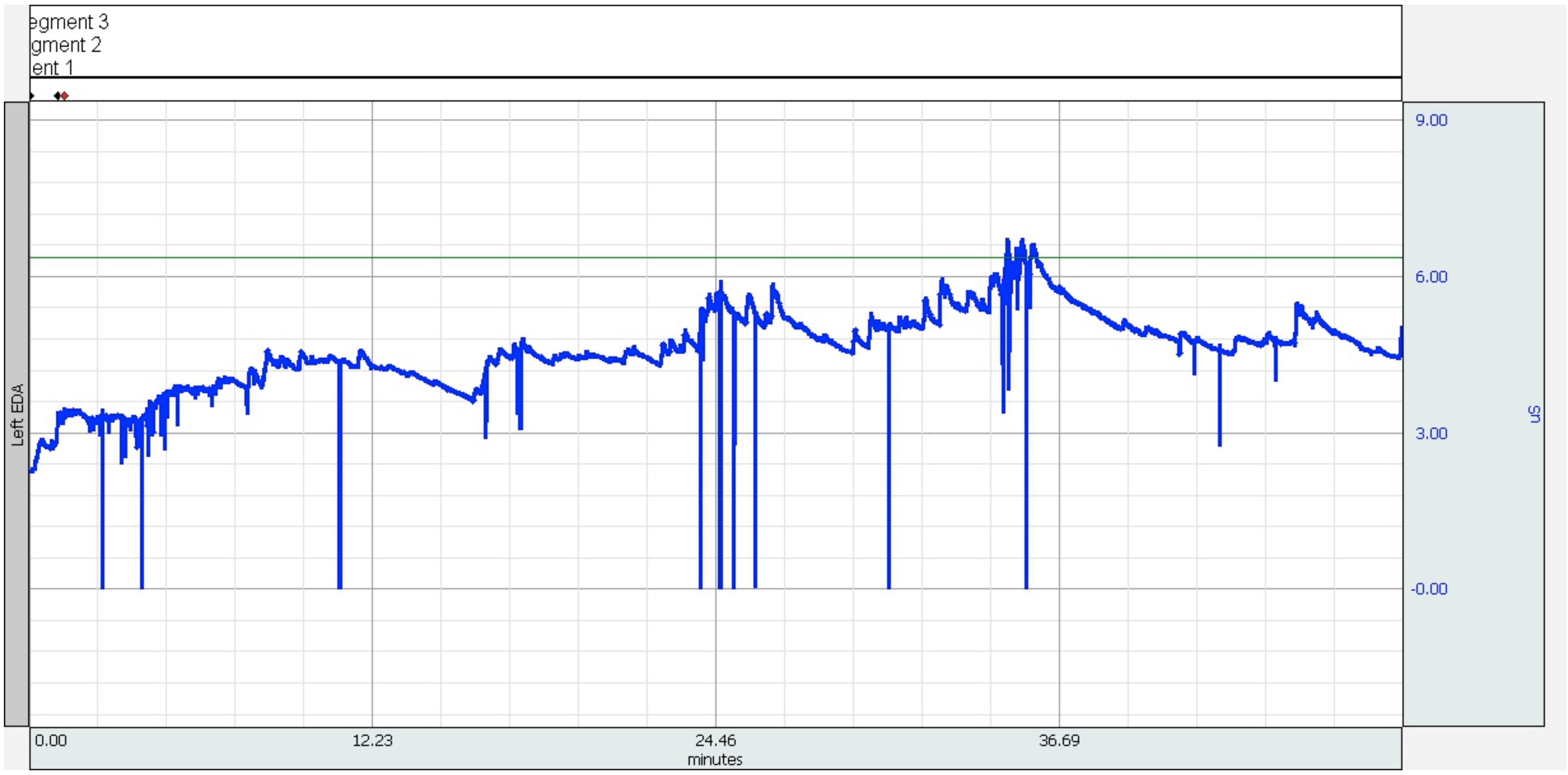
Radix



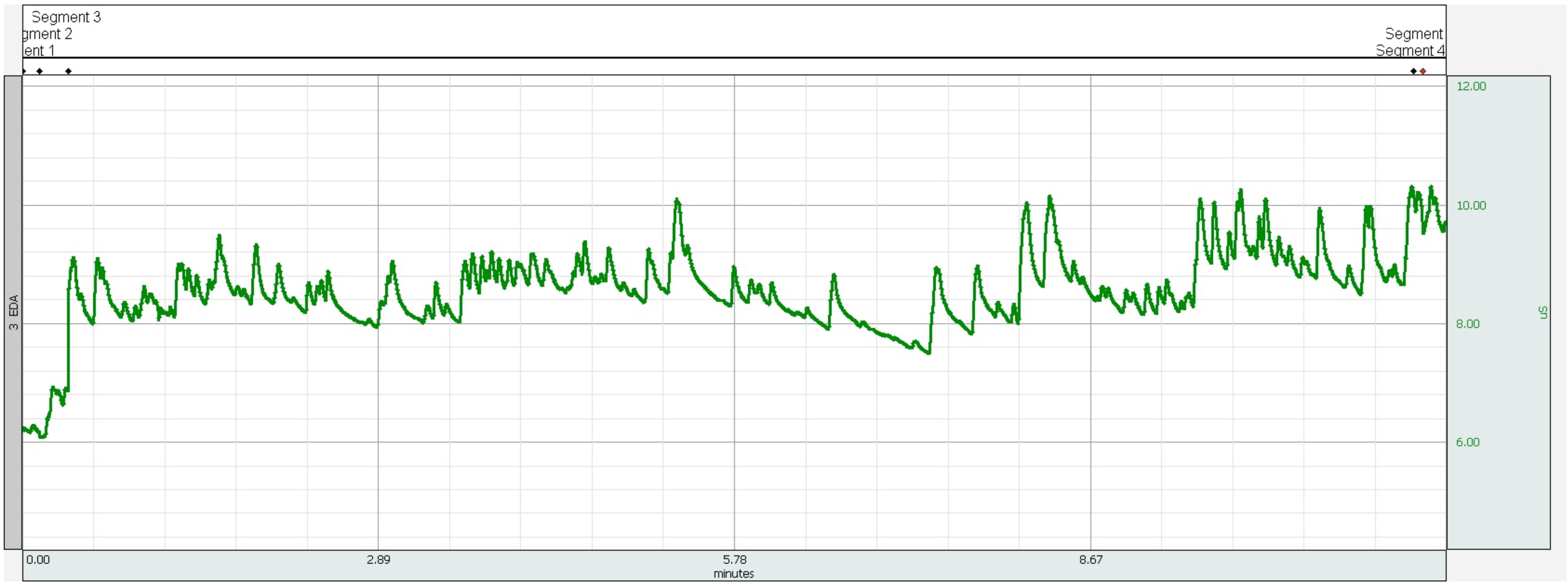
DragonBox

Even though there are small EDA fluctuations in the sampling data of Participant 1, clear interpretative opinions cannot be made. The first participant produces a flat EDA signal that is too insignificant to thoroughly study. As our EDA measuring software is equipped with an interface that automatically scales and zooms, EDA first appears to be responsive on a micro-scale. Only after consulting different sources with regards to overall snapshots of this data, we have concluded that EDA would not register in this case.

Findings: Participant 2

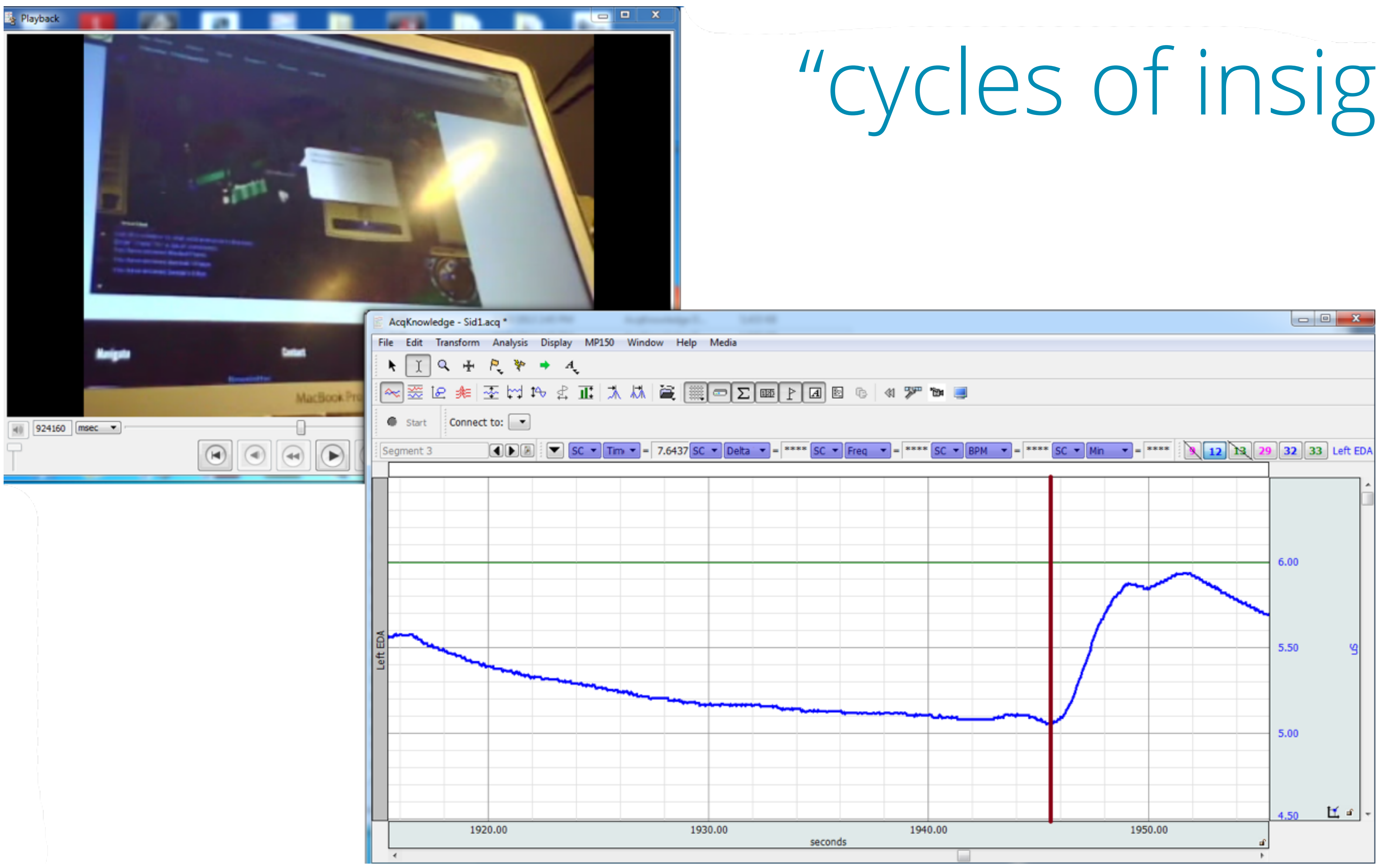


Radix



DragonBox

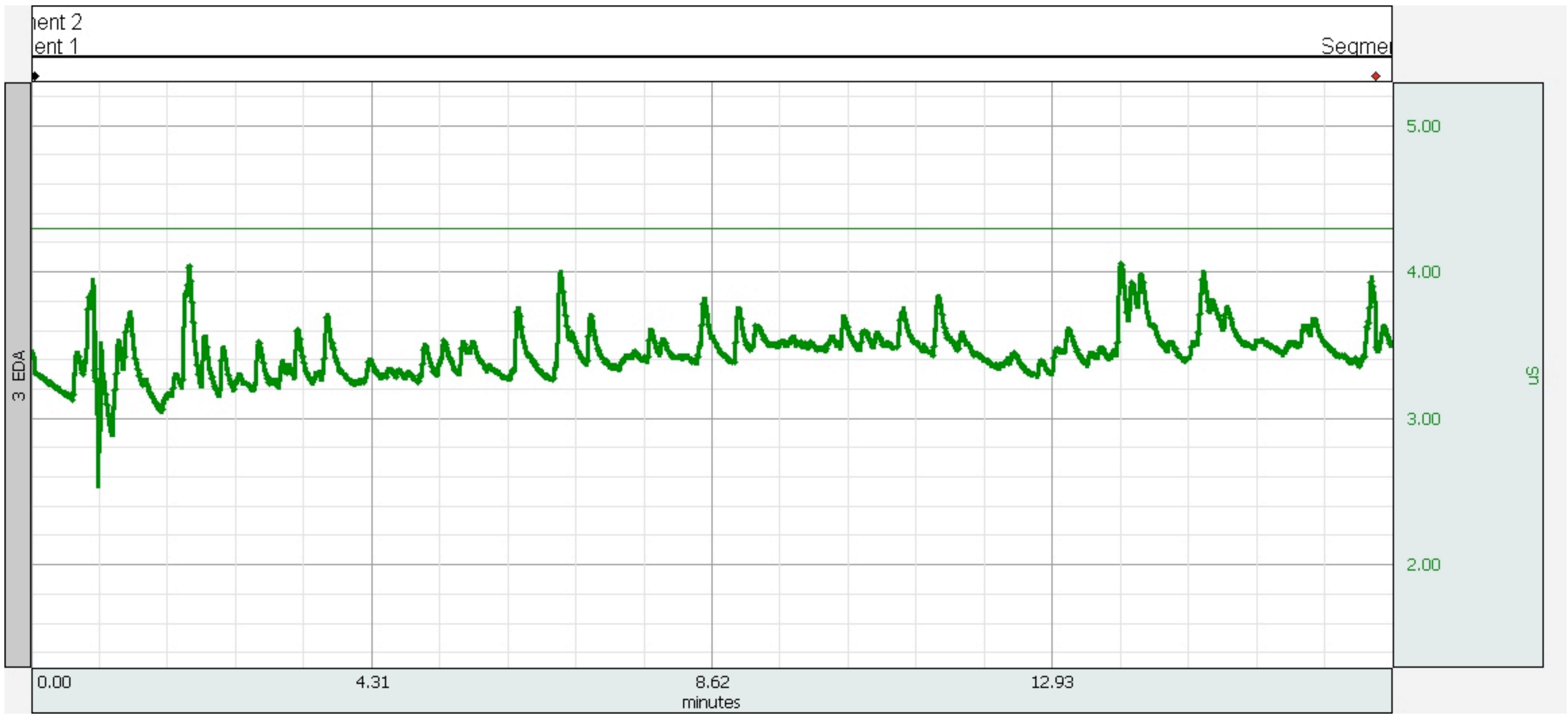
Findings: Participant 2



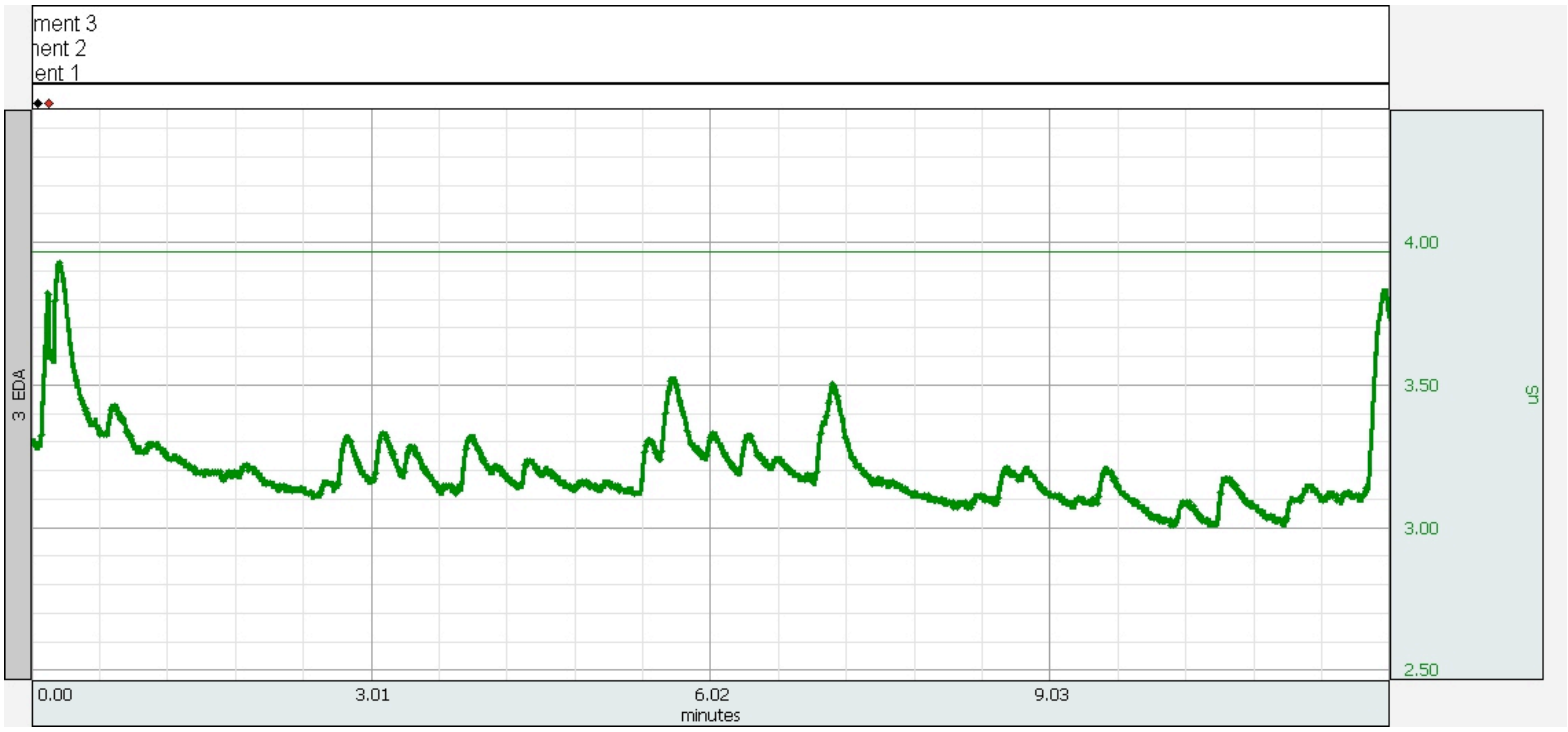
“cycles of insight”

Participant 2 reveals interesting cycles of "insight." In general, the participant is increasingly engaged in Radix via the activation of a patterned repetition of confuse/resolve cycles. Notably, a majority of extended lapses is followed by a peak where the participant reportedly interacts with a new in-game element and gains a new insight. Each insight serves as an EDA stimulant and will help maintaining a high arousal level. Consequently, a higher frequency of insight cycles boosts EDA level while a lower frequency leads to longer lapses and lower arousal.Participant 2 reveals interesting cycles of "insight." In general, the participant is increasingly engaged in Radix via the activation of a patterned repetition of confuse/resolve cycles. Notably, a majority of extended lapses is followed by a peak where the participant reportedly interacts with a new in-game element and gains a new insight. Each insight serves as an EDA stimulant and will help maintaining a high arousal level. Consequently, a higher frequency of insight cycles boosts EDA level while a lower frequency leads to longer lapses and lower arousal.

Findings: Participant 3



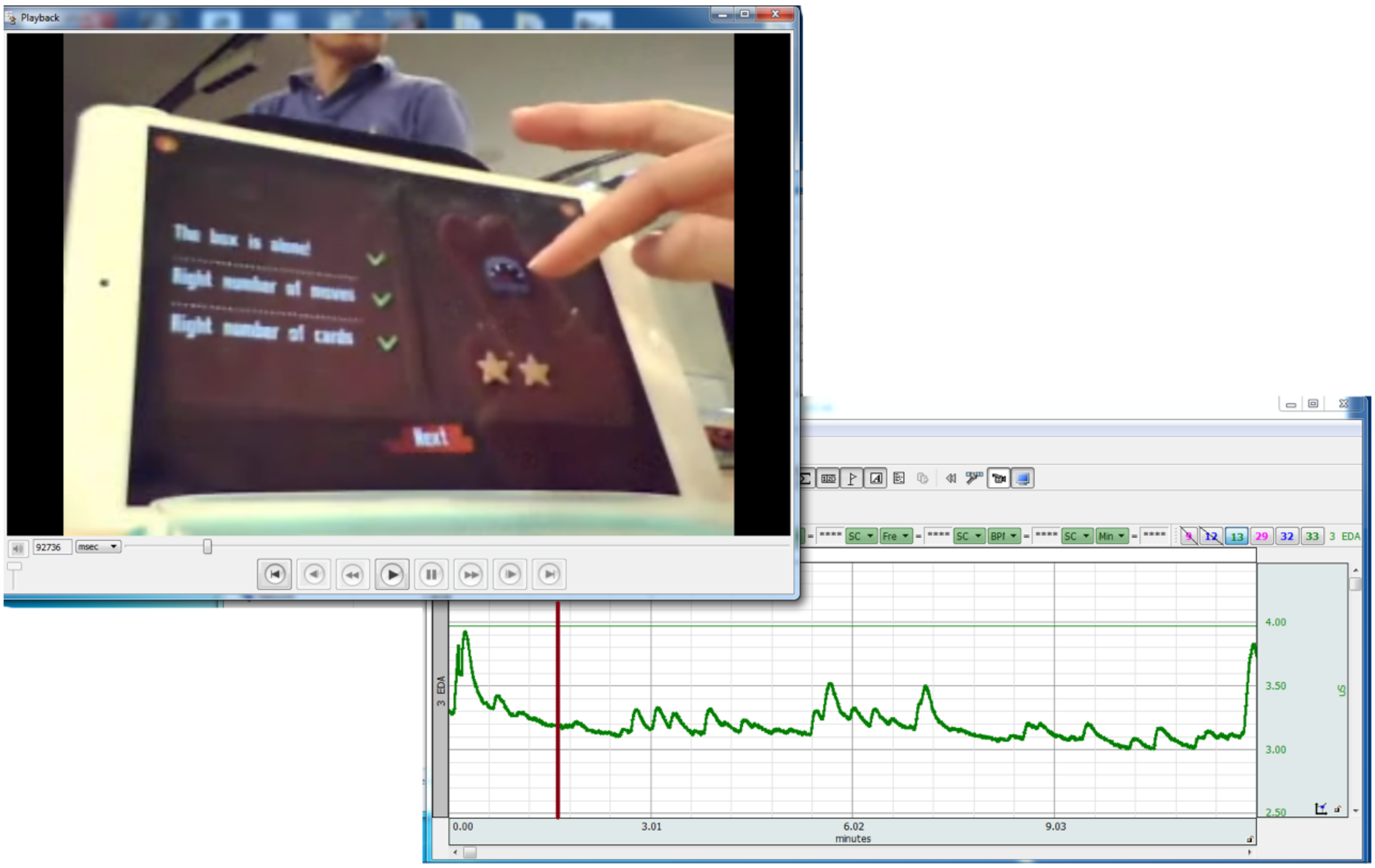
Radix



DragonBox

Findings: Participant 3

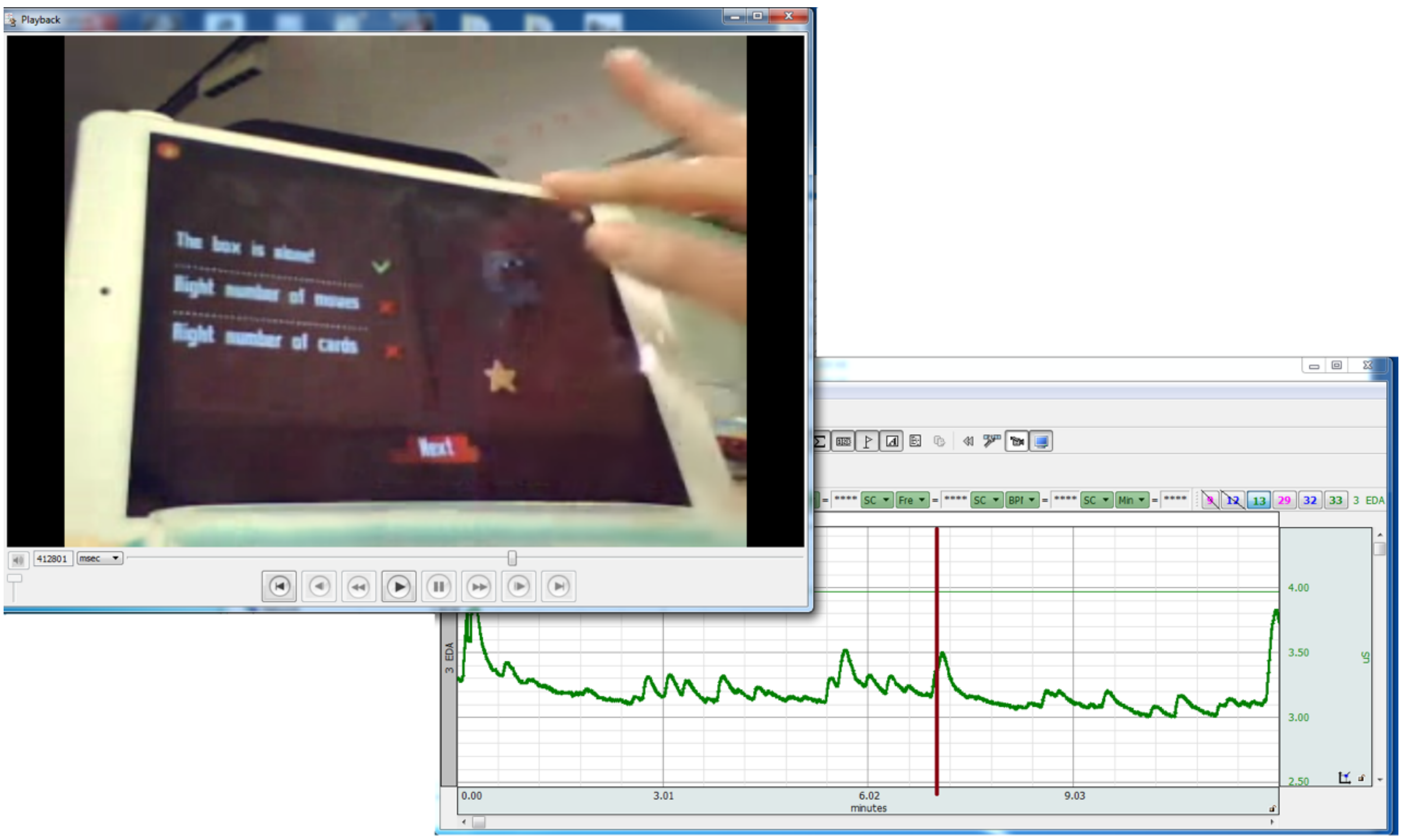
“easy wins versus difficult wins”



Observations of Participant 3 make available further understanding of in-game win/reward mechanism. Throughout Dragon Box playing, we have noticed that the participant inconsistently projects increased EDA at the completion of each level. Upon reviewing gameplay video, we have discovered that, despite the presence of a reward, difficulty plays a key role in shifting one’s arousal. In this first screenshot, an easy win is captured when EDA is solidly declining. In fact, qualitative feedbacks from the participant tell us that the player normally does not feel excited winning a seemingly “easy” level. According to D’Mello and Graesser’s model, this easy-win scenario is lacking a confuse/resolve cycle and is therefore not effective in raising engagement.

Findings: Participant 3

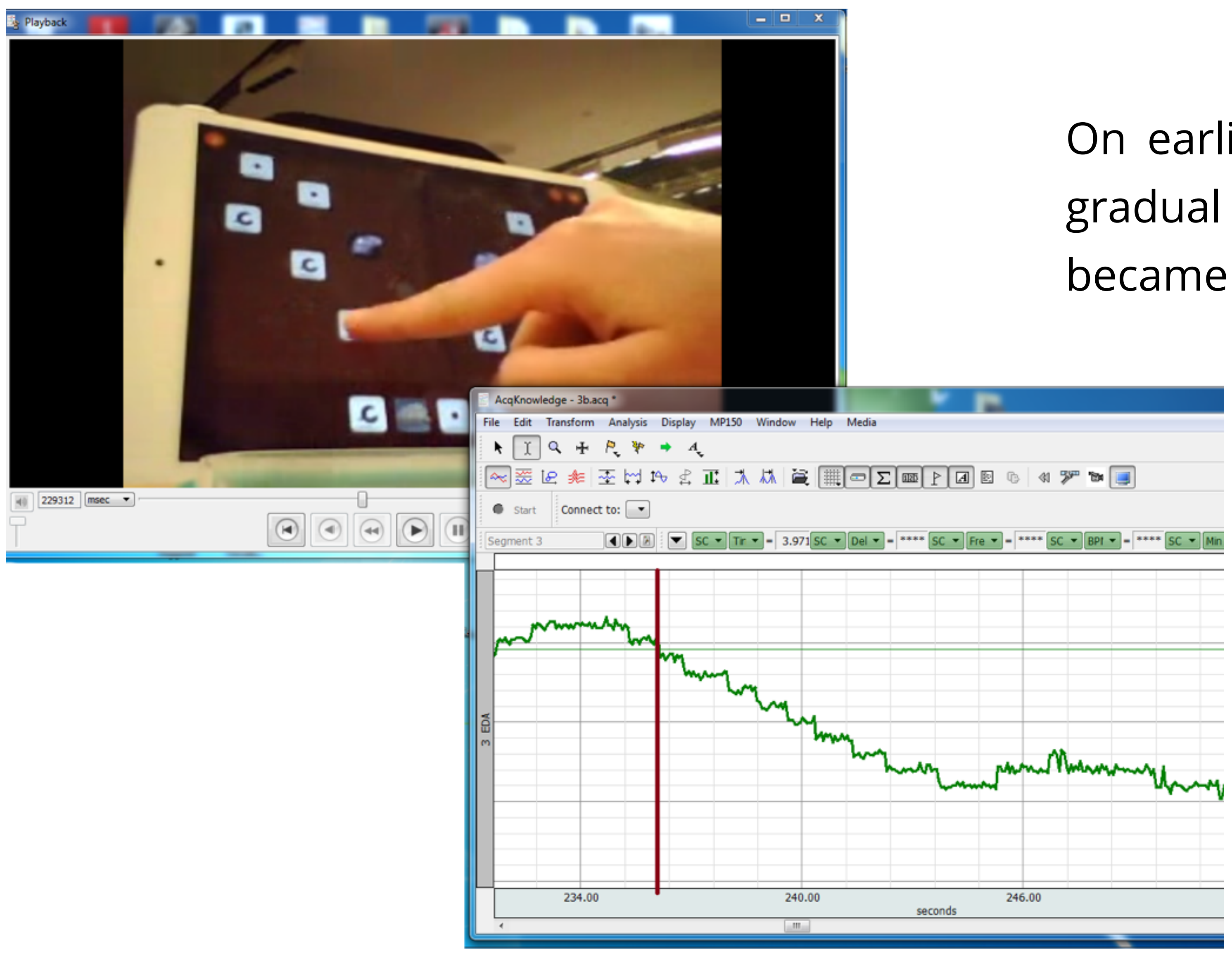
“easy wins versus difficult wins”



Unlike an easy-win, a difficult-win is very predictive of a rise in EDA and arousal. Despite arriving at the same reward screen, a difficult-win is reported to be substantially more exciting. Furthermore, EDA data has shown us that while easy-wins are usually found in lapses of declining EDA, difficult-wins are characterized by patterned data chunks indicating higher-frequencies of multiple confuse/resolve cycles.

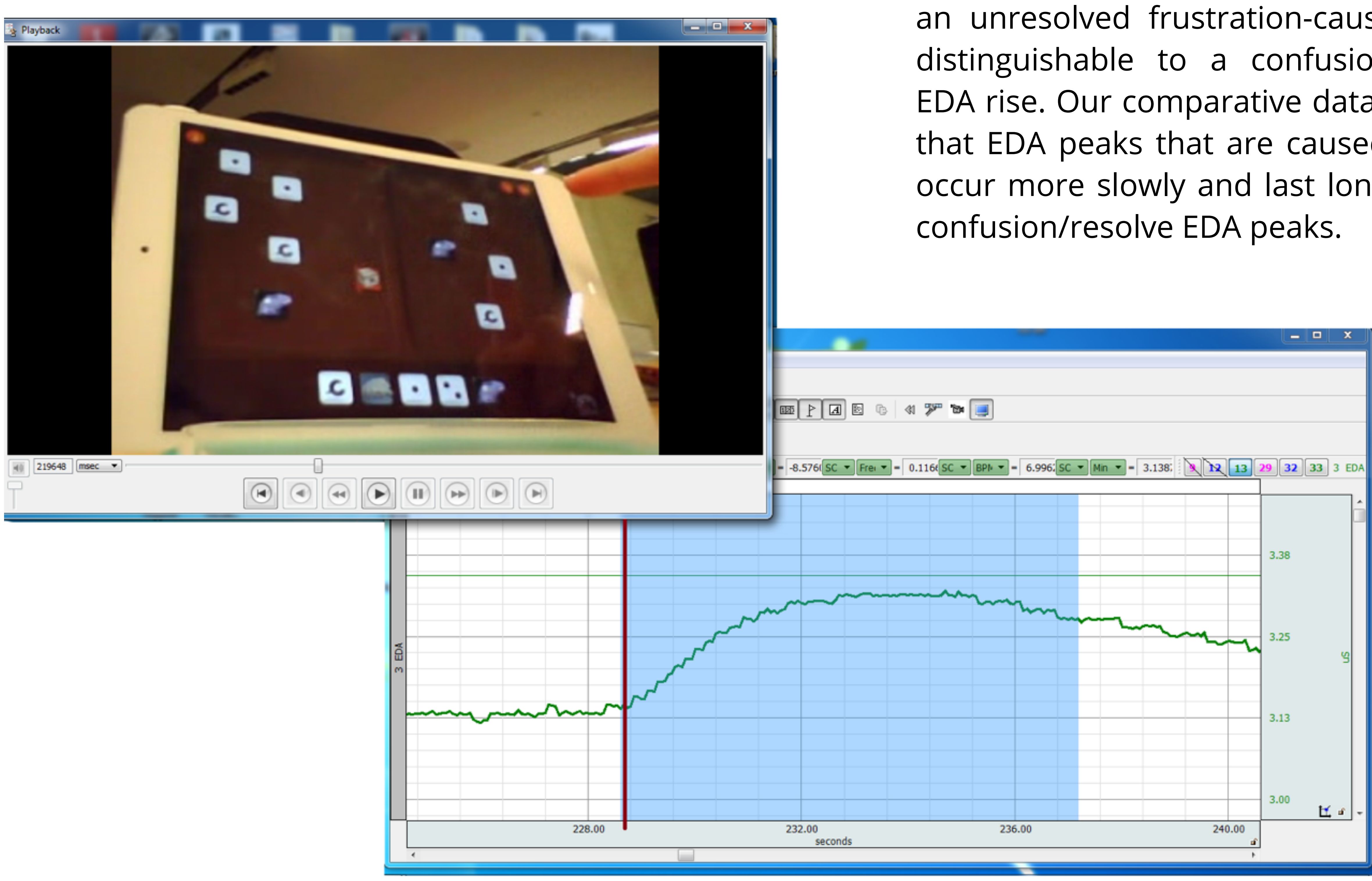
Findings: Participant 3

“playful failure versus frustration”

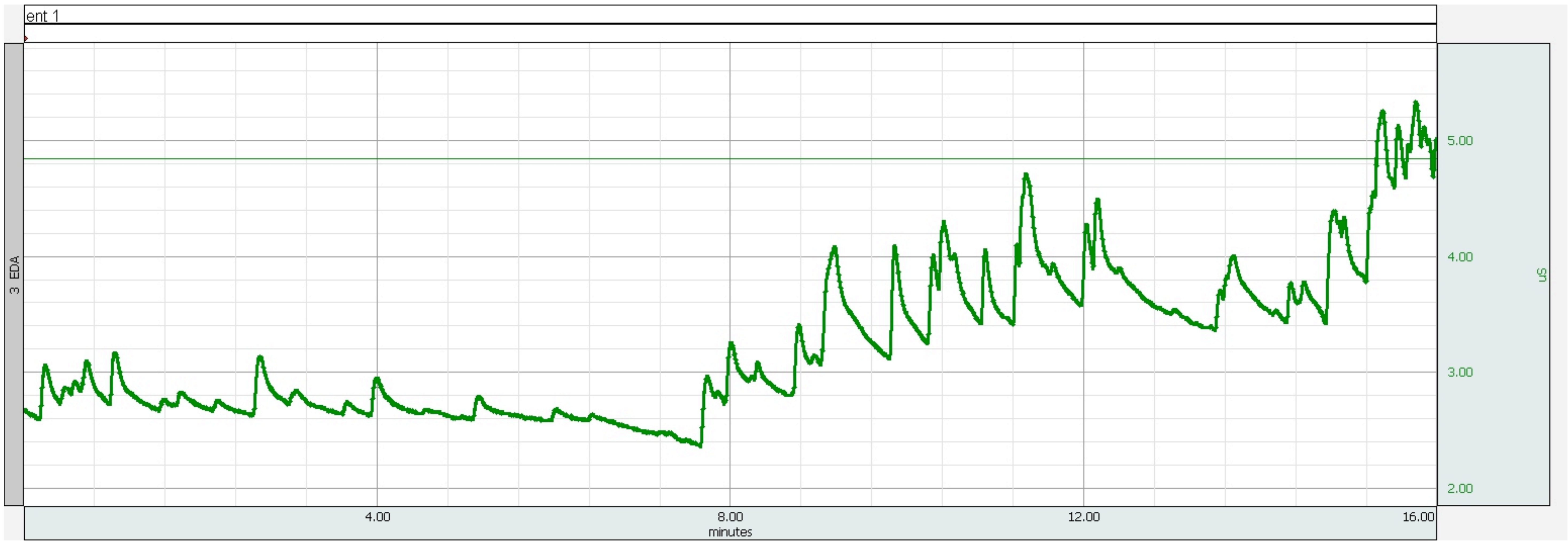


On earlier levels of the game, repeated mistakes produced a gradual decline in EDA. Yet later in the game, as the levels became increasingly difficult, we saw a rise in EDA.

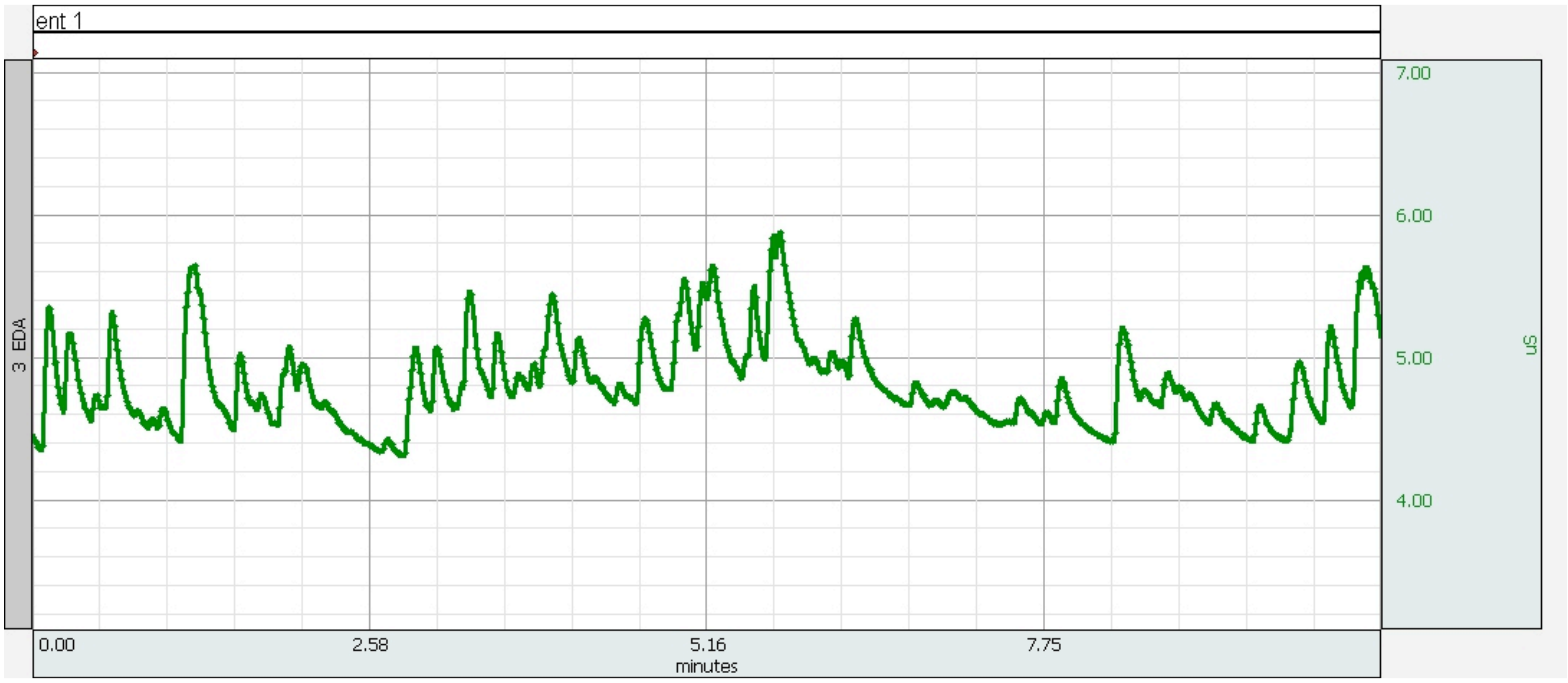
This suggests that the player was less engaged or less concerned with the earlier levels and that perhaps playful failure can possibly transform into frustration. In the second screenshot, Participant 3 spends a large amount of time repeating mistakes and reports becoming frustrated. The point of transformation is marked by a significant rise in EDA, confirming D’Mello’s observation that frustration can lead to arousal. Nevertheless, an unresolved frustration-caused EDA rise is distinguishable to a confusion/resolve cycle EDA rise. Our comparative data analysis shows that EDA peaks that are caused by frustration occur more slowly and last longer than typical confusion/resolve EDA peaks.



Findings: Participant 4



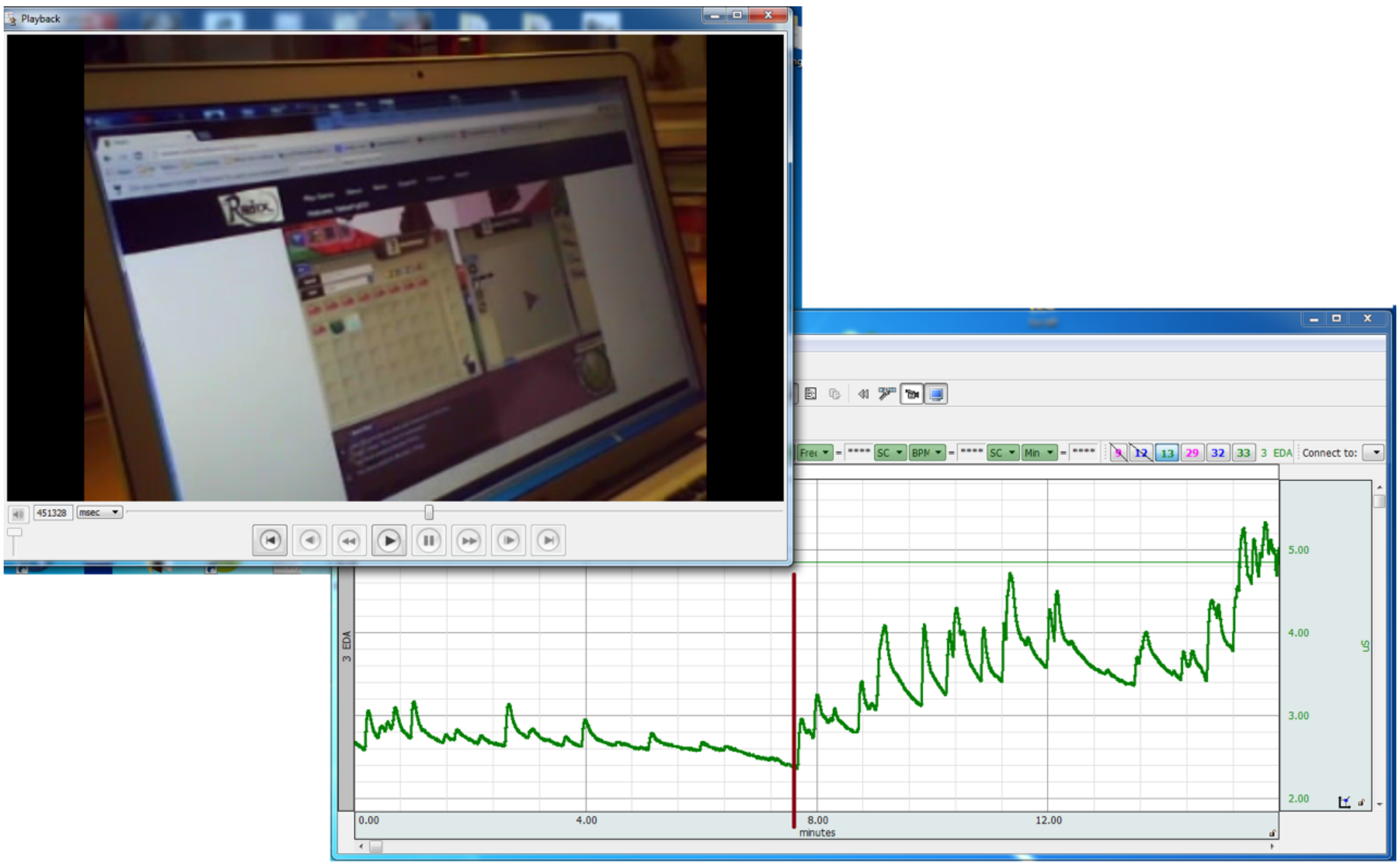
Radix



DragonBox

Findings: Participant 4

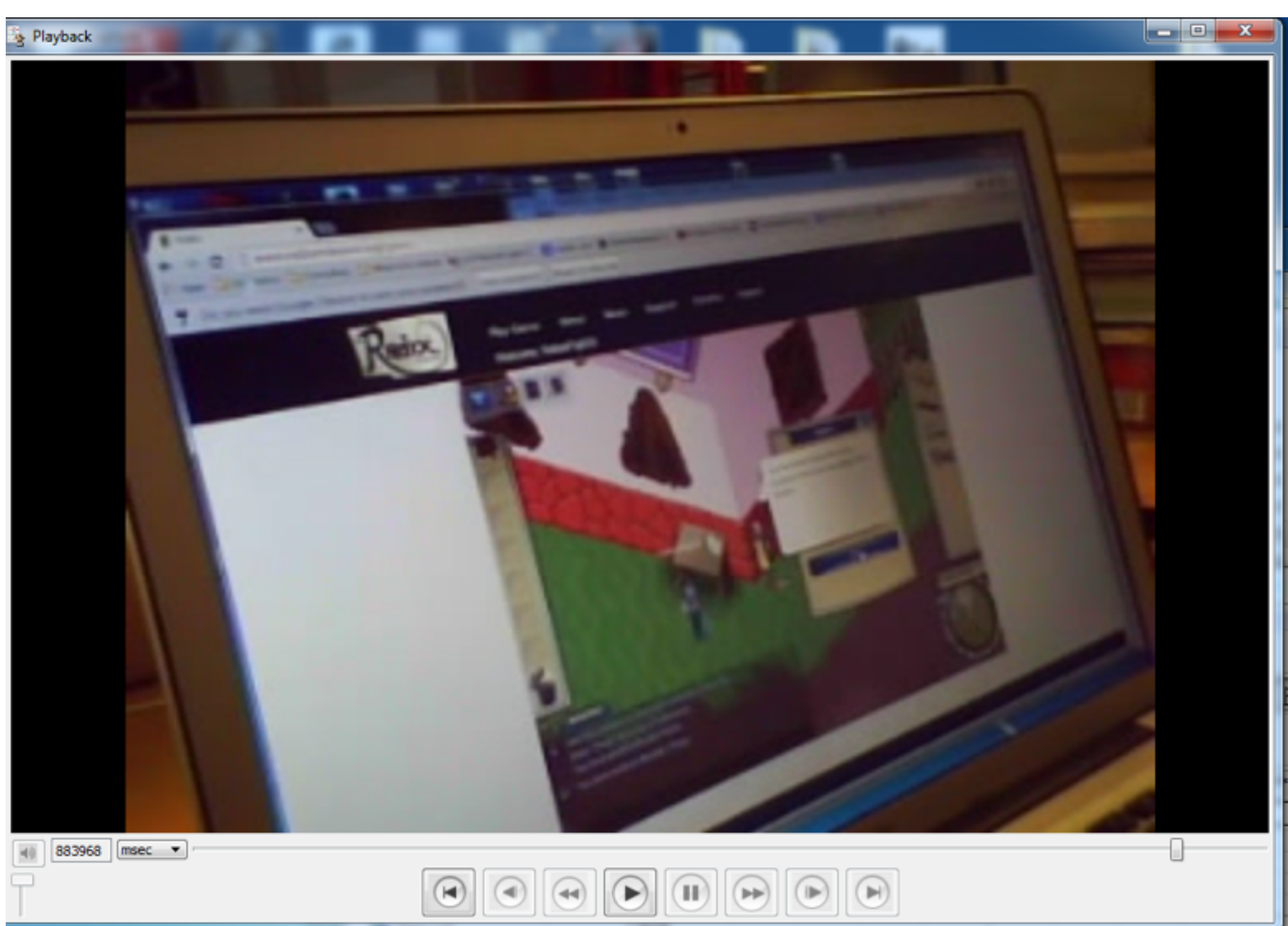
“arousal from game play versus arousal from learning”

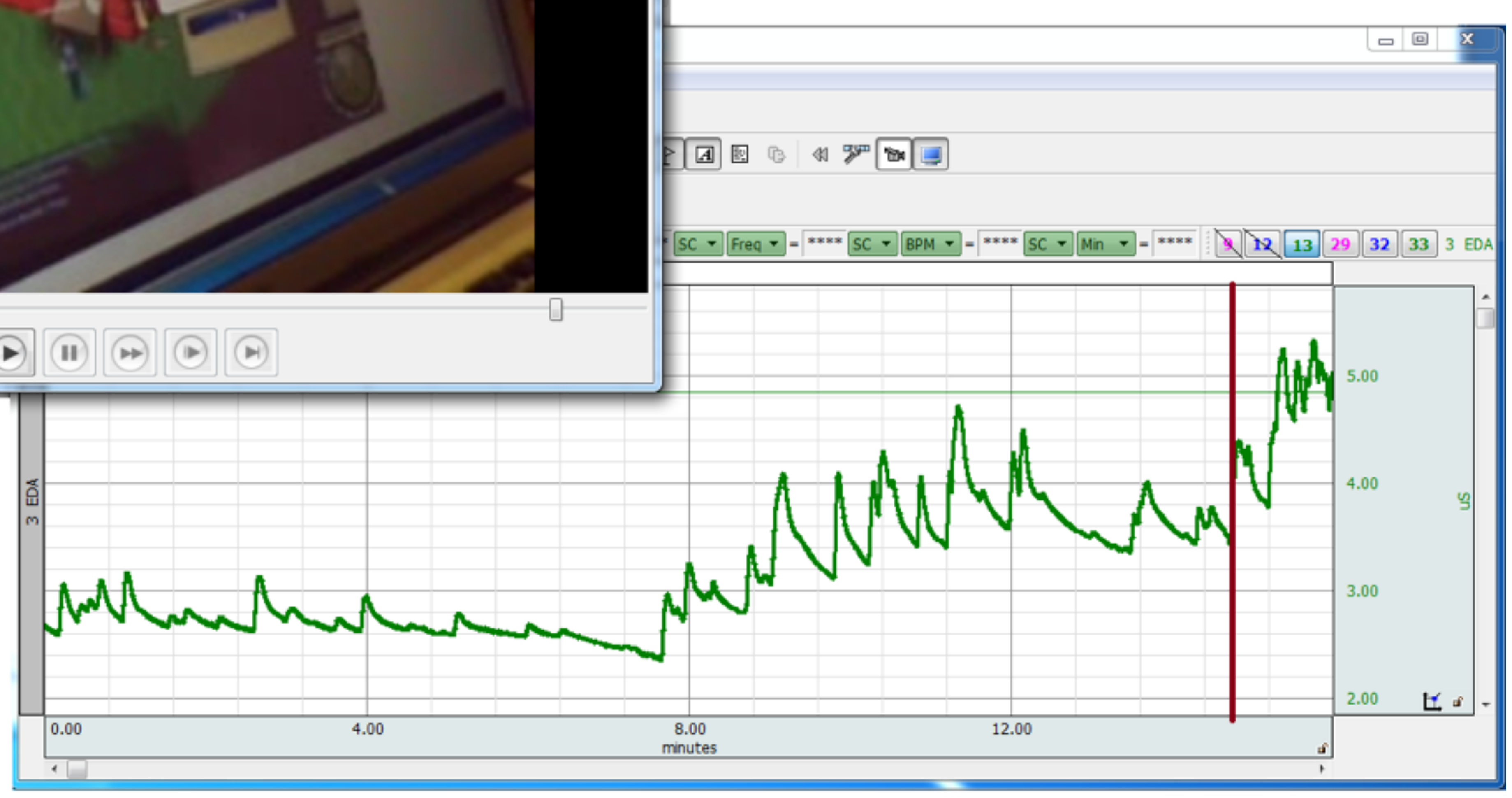


Participant 4 expresses varieties of arousals, resulting from game play versus learning. Namely, game play commonly generates higher arousal than learning. In the first screenshot, Participant 4’s first interaction with a new game mechanism is captured. Evidently, this milestone marks a period of heightened EDA level during which the player actively engages in new game play. Before the trigger-event, Participant 4 reports that s/he is reading textual instructions of quests and is confident that s/he has effectively been absorbing information. While arousal from game play is seen producing higher EDA level than from learning, we are unsure of learning engagement in each situation.

Findings: Participant 4

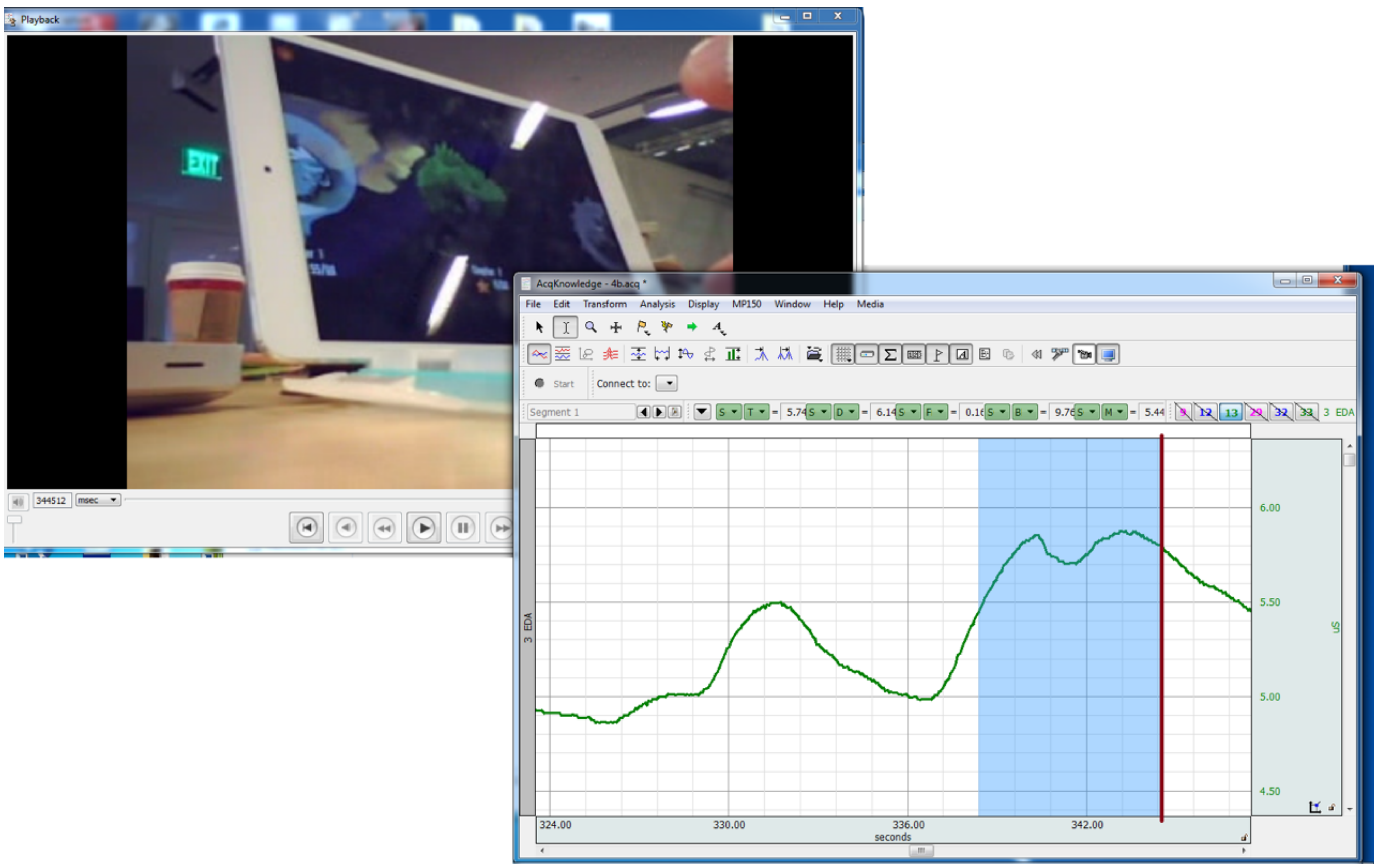
In the screenshot above, we see a slow steady decline as the participant later reported to be unclear how the game could possibly work and began losing interest in it. After the long decline in EDA, the participant continues to play and explore, and we see a sharp increase in EDA when a new NPC (non-playable character) in the game is encountered and offers new insights on how to complete the quest.





Findings: Participant 4

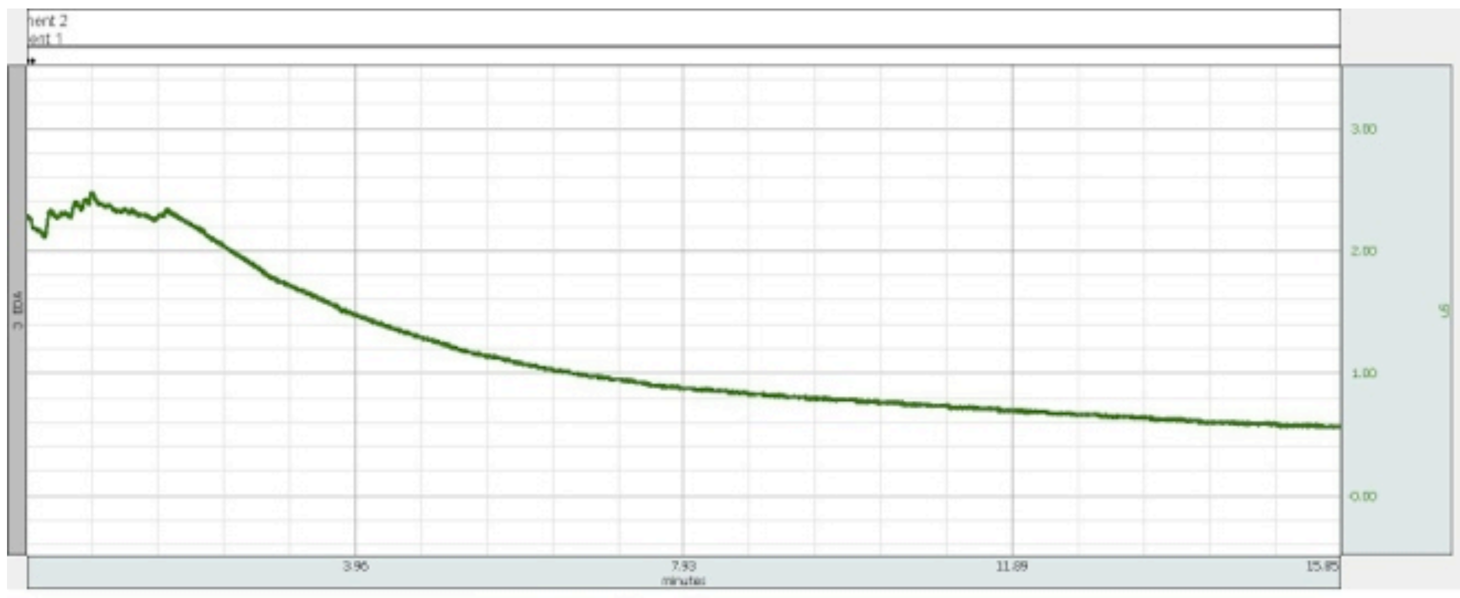
“participant reports versus researcher analysis”



Self-reports were very helpful in supplementing our experiment. For example, Participant 4 demonstrates exhibiting arousal every time a level is finished and is asked to provide an interpretation. According to the player, being presented with a new level and new in-game elements is the main source of excitement. In other words, Participant 4 suggests that the sole appearance of level-completion reward screen is irrelevant. Regardless, upon reviewing EDA data, we find one instance in which Participant 4 finishes his episode and is presented with a level-completion reward screen, followed by an episode-completion reward screen. Explicitly, the latter screen is an indication of total completion and there is no new in-game element is introduced between the two reward screens. However, despite the absence of new in-game elements, two EDA peaks are still seen at the occurrences of both reward screens. This EDA information reveals that, despite the player’s conscious opinions, in-game reward screens successfully register unconscious effect in one’s cognition.

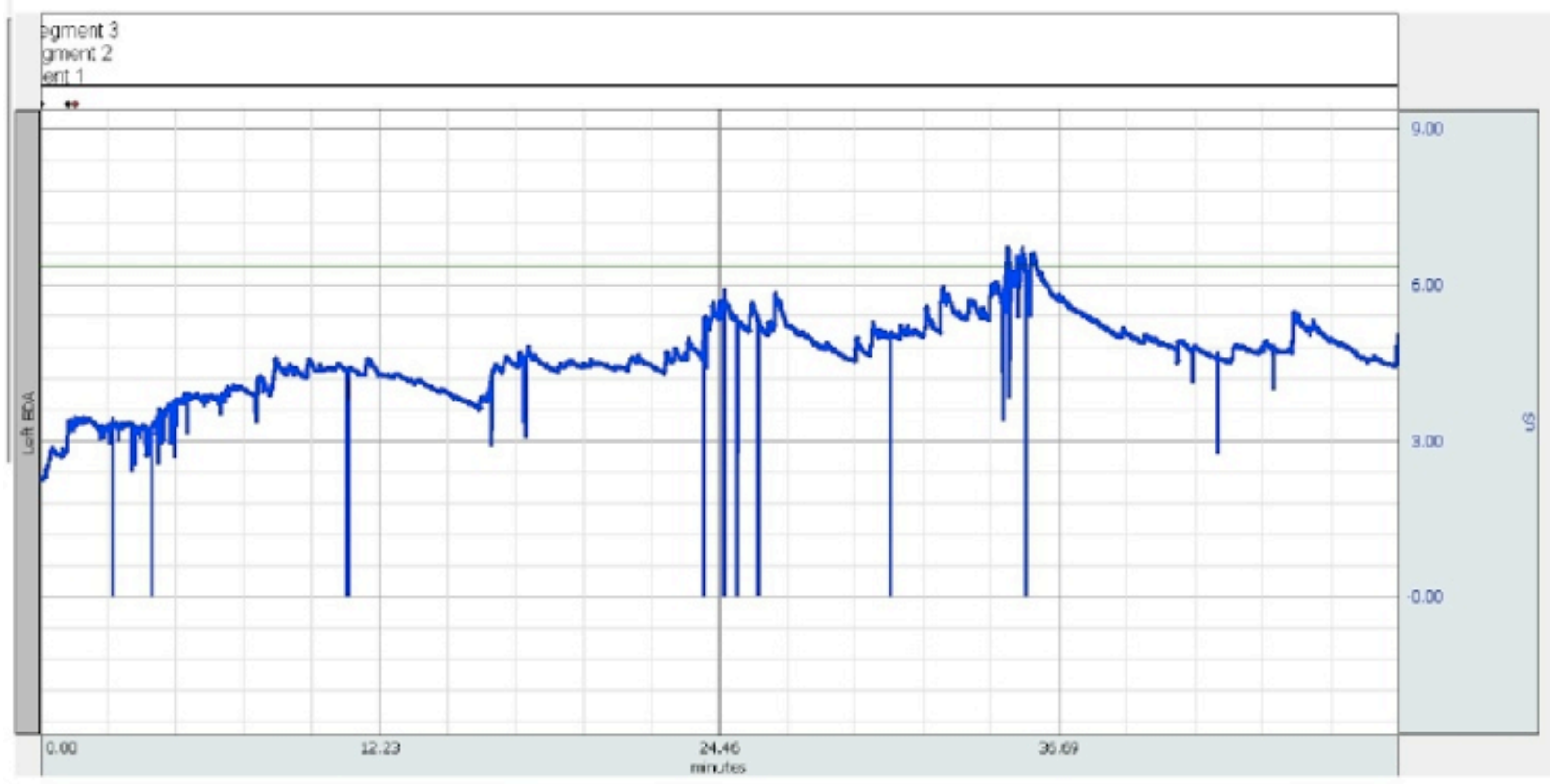
Findings: Overall

- unique data print for each

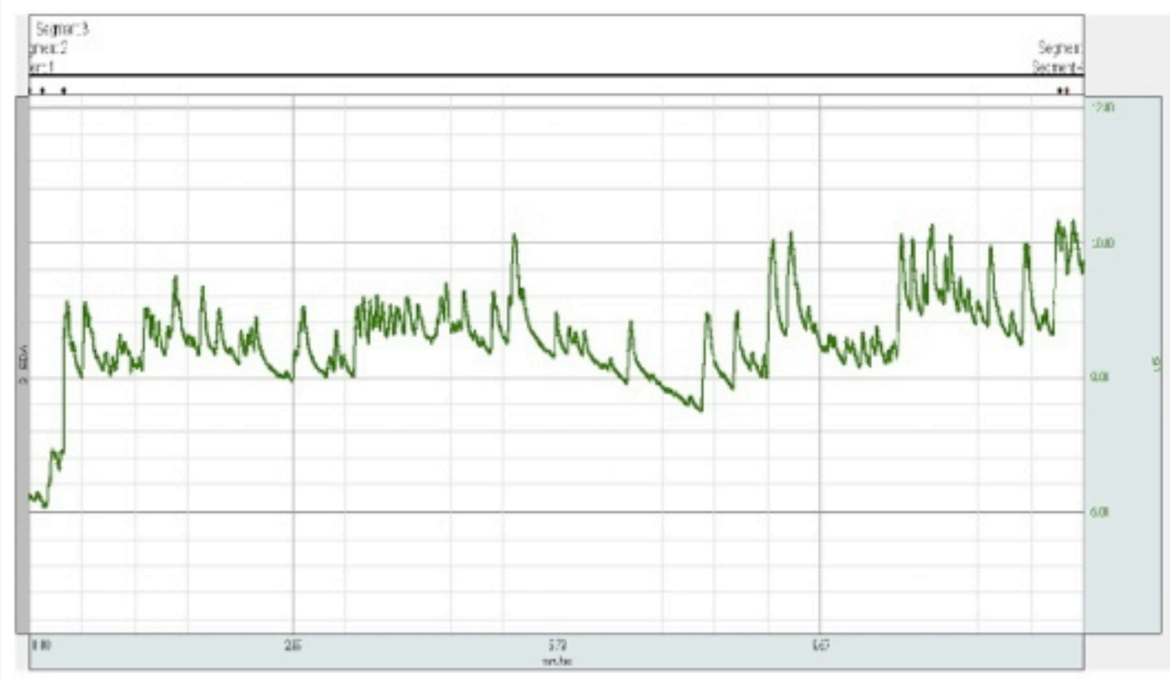


Radix 0

Dragon Box 0

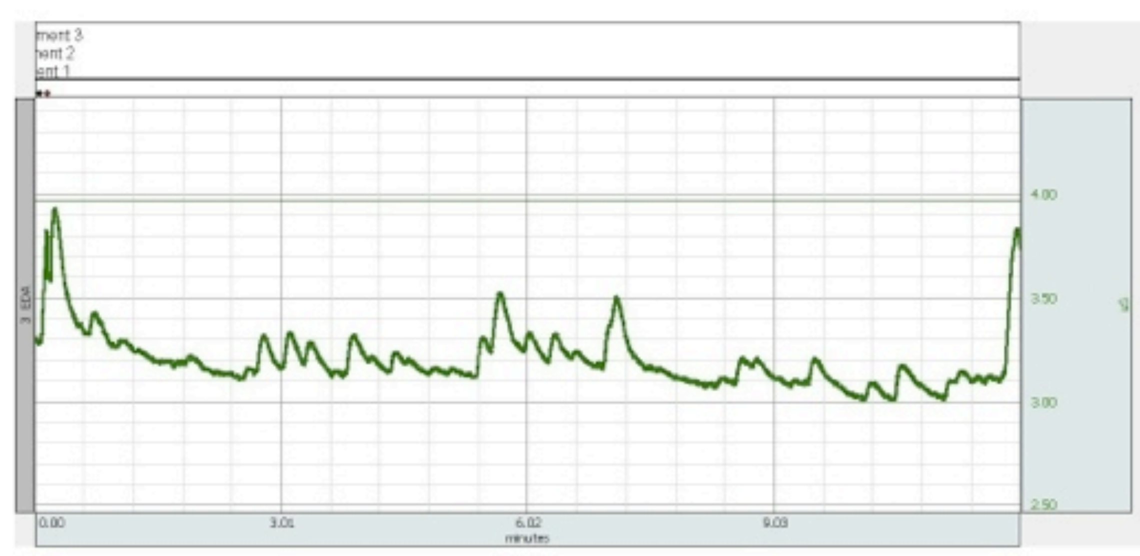


Radix 1

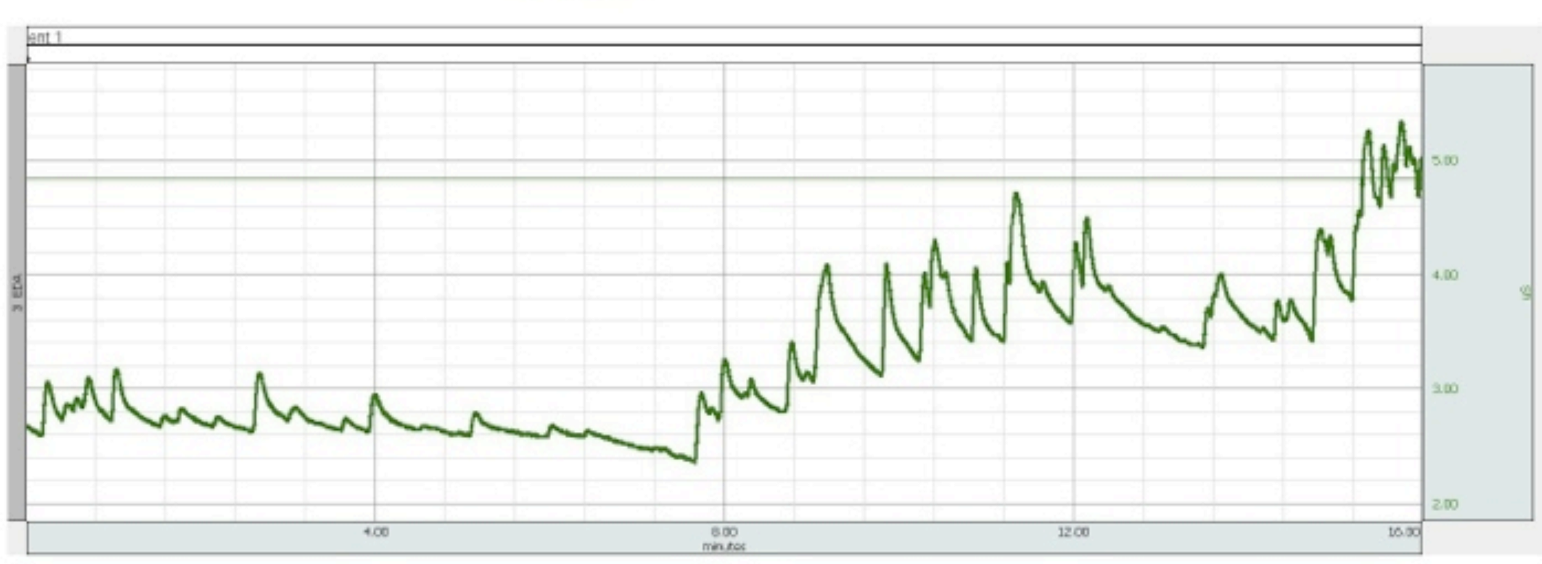


Dragon Box 1

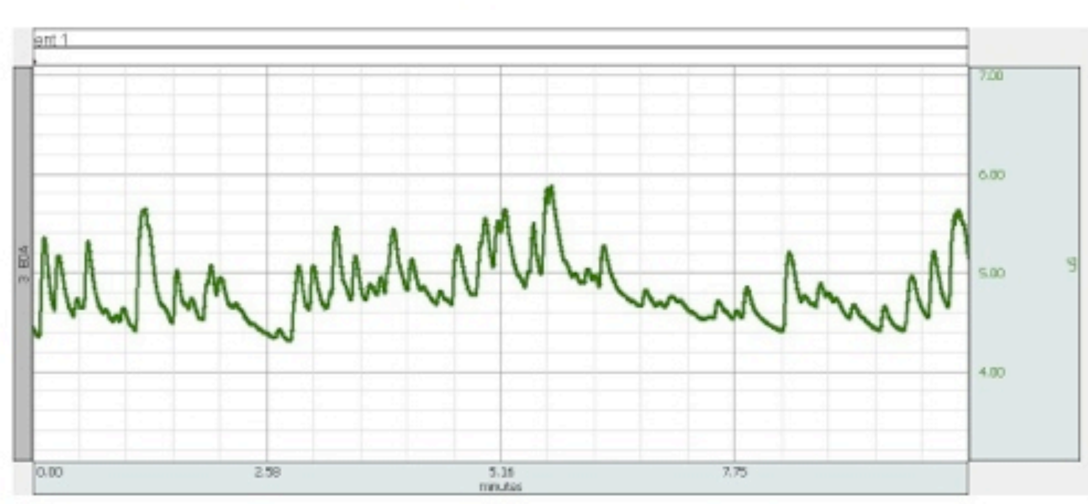
Radix 2



Dragon Box 2



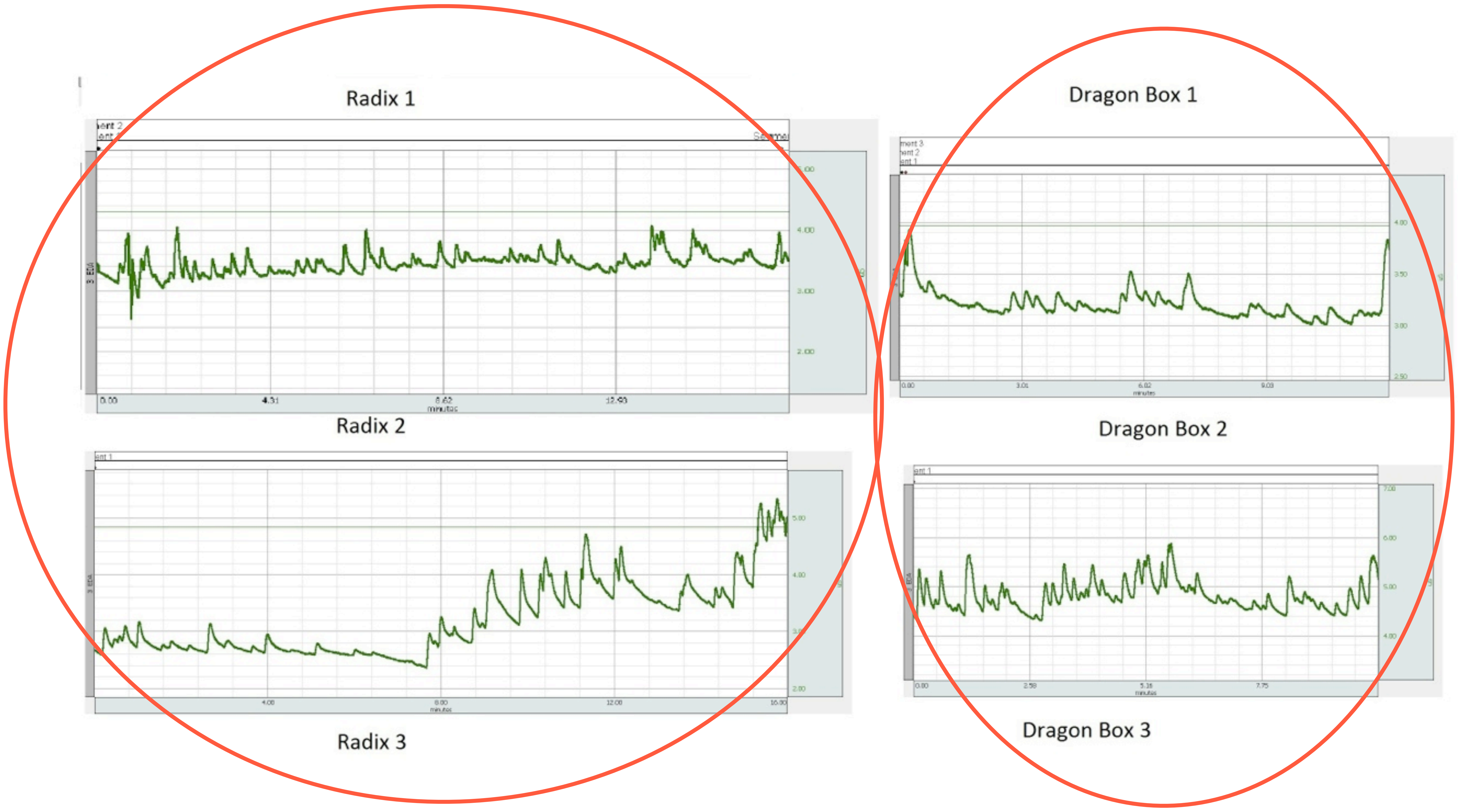
Radix 3



Dragon Box 3

Findings: Overall

- expected and unexpected play patterns



Given the nature of each of the games, we might suspect certain types of EDA patterns for each. Radix being an open-ended exploratory world, we might expect a lower level to start as the player begins to get familiar with the world, the goals of the game, how to navigate, etc., and then as they start to become engaged in the storyline and the tasks at hand, much more EDA activity. We do indeed observe this in Participant 4, but not Participant 3. Yet for Radix, a puzzle game of short bursts of activity in each level, we might expect to see much shorter but frequent cycles of EDA decline and rise as they try a new level and then complete it. We do indeed see this type of activity for both Participant 3 and Participant 4.

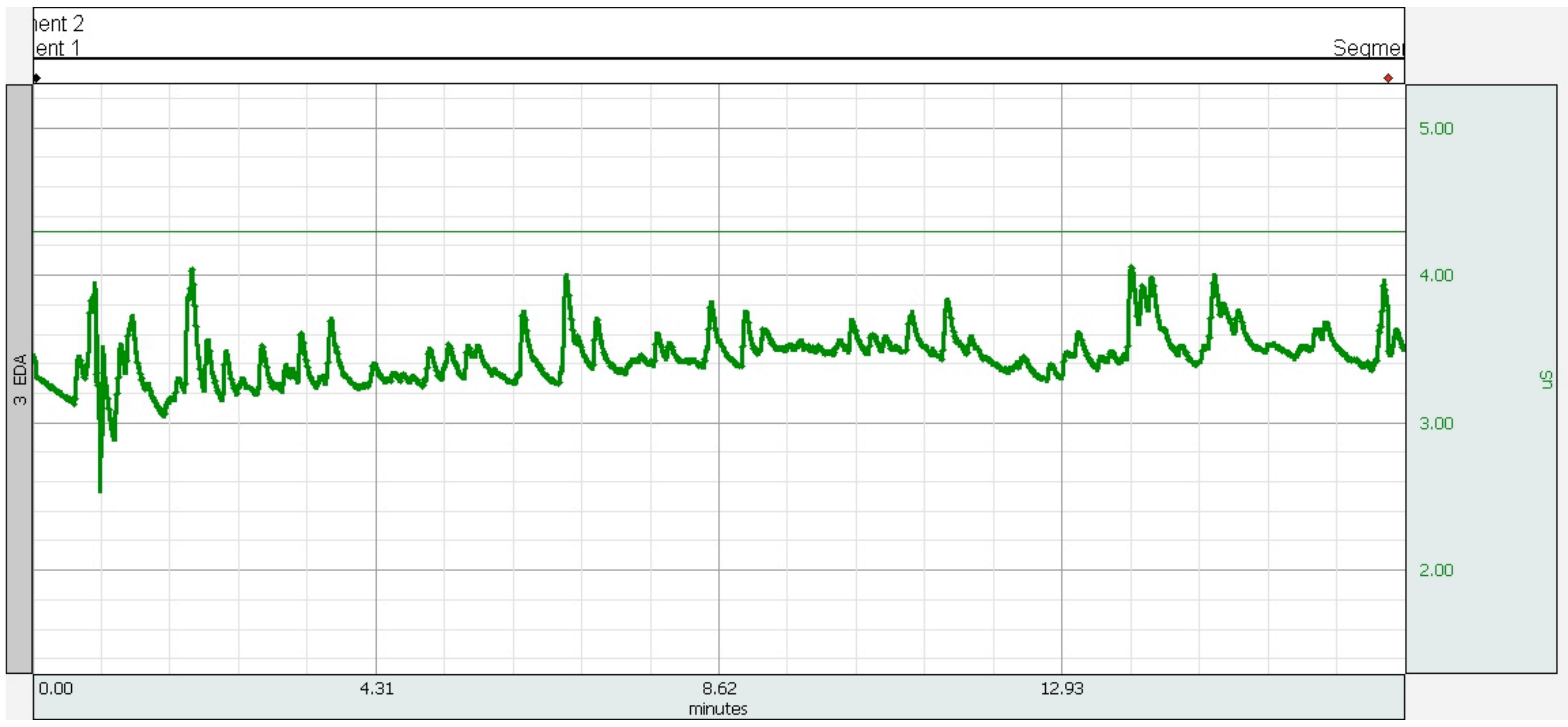
Conclusions

- Participant feedback while reviewing data is very interesting and helpful in making sense of the data.
- Their feedback isn't always consistent with what we see in the data, so we can observe some patterns but there is only so much we can interpret.
- At times we were able to observe small confusion-resolution cycles.
- Player feedback helped us distinguish between game-related confusion and learning-related confusion, but we couldn't determine that from the data alone.
- A more robust data collection mechanism is needed.
- Participants are not always able to provide feedback on the game play.

Takeaways

- we are early in this work, there is much to do yet
- this is a highly valuable mechanism as it is:
    - for learning game designers to get a better sense of how people play their games and the different ways they engage
    - for learners to be more reflective of their own engagement and learning dynamics
    - for curriculum developers and instructional designers to have a more nuanced lens on how learners vary and engage with learning material

Additional Observations



A Connection to Resiliency and Growth Mindset?
An interesting observation was participant 3's feedback on their experience with Game 1, Radix. She gave fairly negative feedback about the game, but largely in relation to it's mechanics—statements such as, "the game mechanics were confusing", "it wasn't clear what I had to do, I was just wandering around the world", etc. We also noticed that in the video of her game play she abandoned techniques quickly and jumped between puzzles quickly. This was also the EDA dataset that seemed very flat, as though she had never engaged with the data—and that this genre of game was very different than the types of games she said she often played.

Though we certainly can make concrete inferences around this, it did remind of us portion of D'Mello et al. paper, "Confusion can be beneficial for learning" where they discuss confusion induction as a technique for creating the opportunity for deeper learning:

> *"There is also the manner of identifying who might benefit from a confusion induction intervention. It is probably not a very sensible strategy to attempt to confuse a struggling learner or to induce confusion during high stakes learning activities, at least until confusion induction techniques are refined and their consequences are better understood. Currently, these interventions are ideally suited for gifted learners who are often bored and disengage from learning activities due to a lack of challenge (Pekrun, Goetz, Daniels, Stupnisky, & Perry, 2010). There is also a risk of confusing students who are cautious learners instead of academic risk takers (Clifford, 1988; Meyer & Turner, 2006) or learners who have a fixed (entity) instead of growth (incremental) mindset of intelligence (Dweck, 2006). Confusion interventions are best for adventurous learners who want to be challenged with difficult tasks, are willing to risk failure, and manage negative emotions when they occur because they understand that failure is an inevitable part of a successful path towards proficiency development. These learners can be challenged at the extremes of their zones of proximal development (Brown, Ellery, & Campione, 1998; Vygotsky, 1986), provided that appropriate scaffolds are in place when they struggle or they can manage the challenges with self-regulated learning strategies."*

~D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A., n.d., p. 47

This is very interesting to us indeed, as Carol Dweck's work on *mindset* (see Dweck, 2006) and Angela Duckworth's work on *resiliency/grit* (see Duckworth et al, 2007) – and the impact of both on learning outcomes – is at the forefront of the fields of education and learning sciences. The participant's feedback and our observation of her game play did suggest that her engagement in the game was low and her willingness to persist through exploring how to play the game, what do next, etc., was also low. While indeed there could be a number of factors contributing to this, including general disinterest in this game alone, in the study, etc., it did leave us very interested in further exploration of how affective computing tools might be able to better identify and measure such critical but difficult to access traits as *mindset* and *resiliency/grit.*

Other Things We Learned

- It was difficult to get some participants to register for skin conductance/EDA
- there's a significant learning curve with the tools
- you have to situate mobile games so that you don't interfere with the EDA

References

Azavedo, R., Bouchet, F., Harley, J., Feyzi-Behnagh, R., Trevors, G., Duffy, M., Taub, M., Pacampara, N., Agnew, L., Griscom, S., Mudrick, N., Stead, V. & Yang, W.(n.d.). MetaTutor: An intelligent multi-agent tutoring system designed to detect, track, model, and foster self-regulated learning.

Craig, S., Graesser, A., Sullins, J., & Gholson, J. (2004). Affect and learning: An exploratory look into the role of affect in learning. Journal of Educational Media, 29, 241-250.

Csikszentmihalyi, M., & Csikszentmihalyi, I. S. (1988). Optimal experience: Psychological studies of flow in consciousness. Cambridge, England: Cambridge University Press.

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. Learning & Instruction, 29, 153-170.

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (n.d.). Confusion Can Be Beneficial for Learning. A report for the National Science Foundation.

D'Mello, S. & Graesser, A. (2012). Dynamics of affective states during complex learning. Learning & Instruction, 22(2), 145-157.

Dweck, C. (2006). Mindset: The new psychology of success. Random House Digital, Inc..

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. Journal of Personality and Social Psychology, 92(6), 1087-1101.

Graesser, A. & D'Mello, S. (2011). Theoretical perspectives on affect and deep learning. In, Calvo, R. & D'Mello, S. (eds.) New Perspectives on Affect and Learning Technologies, 11-22.

Graesser, A. C., D'Mello, S. K., Chipman, P., King, B., & McDaniel, B. (2007). Exploring relationships between affect and learning with AutoTutor. In R. Luckin, K. Koedinger, & J. Greer (Eds.), Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work, 16–23. Amsterdam: IOS Press.

Graesser, A. C., Jackson, G. T., & McDaniel, B. (2007). AutoTutor holds conversations with learners that are responsive to their cognitive and emotional states. Educational Technology, 47, 19–22.

Kiili, K. & Lainema, T. (2008). Foundation for measuring engagement in educational games. Journal of Interactive Learning Research, 19(3), 469-488.

Lehman, B., D'Mello, S., & Graesser, A. (unpublished manuscript). Interventions to regulate confusion during learning.

Strike, K. & Posner, G. (1992). A Revisionist Theory of Conceptual Change. In R. A. Duschl & R. J. Hamilton (Eds.) Philosophy of Science, Cognitive Psychology, and Educational Theory and Practice, 147-175. New York: State University of New York Press.

Work Statement

Below are the tasks completed by the researchers:

- COUHES form (Jen)
- Literature review for initial presentation (Trung)
- Game identification and selection (Jen)
- Facial affect recognition software investigation (Trung)
- EDA tool investigation and setup (both)
- Participant recruitment (Jen)
- Running participant sessions (both)
- Literature review for project setup and final submission (Jen)
- Data analysis (both)
- Photographs and aggregated snapshots (Trung)
- Final presentation development (both)
- Final submission development (both)