Towards an Optimal Affect-Sensitive Instructional System of Cognitive Skills

Jacob Whitehill
Machine Perception Laboratory
University of California, San Diego (UCSD)

in collaboration with

Zewelanji Serpell, Aysha Foster, Yi-Ching Lin, Brittney Pearson, Marian Bartlett, and Javier Movellan

UCSD

Temporal Dynamics of Learning Center
An NSF Science of Learning Center

Machine Perception Lab

- Research goals:
  1. Develop automated tools to perceive faces, facial expression, and emotion
  2. Study natural human behavior using computational methods.
  3. Develop intelligent systems and social robots that interact with humans autonomously.
     - Automated teaching systems

Affect-sensitive teachers

- We are interested in the role of **affect** in teaching and automated teaching systems.
- What does it take to build an affect-sensitive automated teacher that is more effective than a comparable affect-blind system.
    - For what kind of learning domains does affect-sensitivity make a difference?
    - How can we detect the key affective states of the student?
    - How can affective state estimates be integrated into the decision engine?

Affect-sensitive teachers

- Affect:
  - Emotional & motivational state of student
  - In practice: any information useful for teaching that can be gained from video of the student
- Affect-sensitive automated teacher:
  - Teaching system that models/senses student's affective state to make decisions.
  - Contrast with "affect-blind" teacher.

Affect-sensitive teaching

- Is affect-sensitivity really necessary for good teaching?
  - Many teachers say that they use students' facial expressions to modulate how they teach.
  - However, empirically there is little evidence to show that affect-sensitivity is really necessary or yields any learning gains.

Affect-sensitive teaching

- D'Mello, Graesser, et al. 2010 conducted 2-day study of affect-sensitive versus affect-blind computer literacy tutor.
  - Day 1: Affect-sensitive computer literacy tutor was statistically significantly *less* effective than affect-blind system on day 1.
  - Day 2: Affect-sensitive tutor was slightly (not stat. sig.) more effective than affect-blind system.

Why might affect be important?

1. Some affective states may be **more desirable** than others.
   - It may be undesirable for a student to become very frustrated or upset.
2. Modeling affect and utilizing affective sensors can help to **disambiguate states**.
   - If a student's performance drops suddenly, is it because the task is too hard, or because he/she is not trying?
     Should the teacher make the task harder or easier?

Why might affect be important?

3. "Affective sensors" may provide useful information about student state at a **finer timescale** than is possible from students' explicit input (keyboard presses, mouse clicks, touches, etc.).

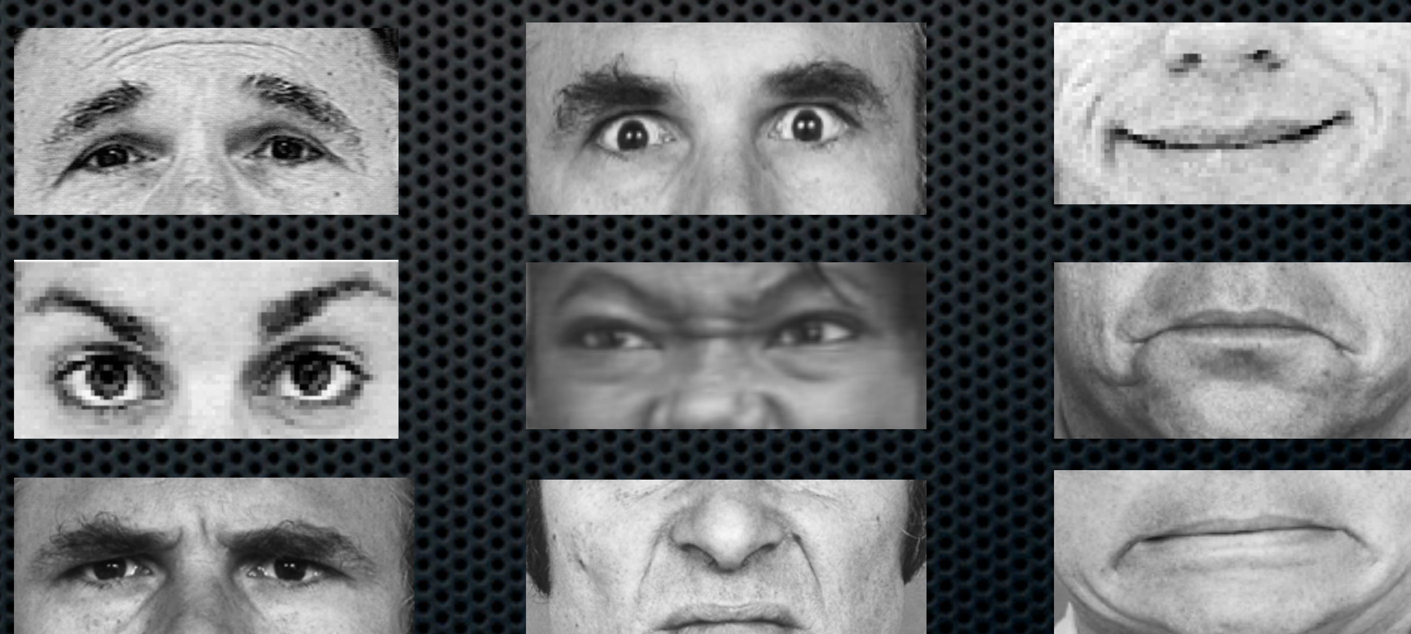
   - Can we predict if the curriculum is too hard/easy for the student in real time, before he/she makes a mistake or gets bored?

Computer Expression Recognition Toolbox (CERT)

- Over the past 15 years we have developed an array of tools for fully automatic, real-time face analysis.
  - Together, these tools form the Computer Expression Recognition Toolbox (**CERT**).
- CERT facilitates analyses of human behavior at finer timescale than possible with human coding.
- CERT also provides a real-time feedback signal to intelligent systems.

Computer Expression Recognition Toolbox (CERT)

- CERT is a fully automatic real-time system for face processing.
  - Basic emotion recognition
  - Facial action recognition (FACS, Ekman & Friesen 1978)
  - Facial feature positions (eyes, nose, and mouth)
  - 3-D head pose estimation



Examples of facial actions

Computer Expression Recognition Toolbox (CERT)

- CERT is available for free academic use: see http://mpt4u.com
- (1 min demo)

Computer Expression Recognition Toolbox (CERT)

- In 2008 we performed a pilot study:
    - Can automatic facial expression recognition be used to estimate a student's perception of difficulty?

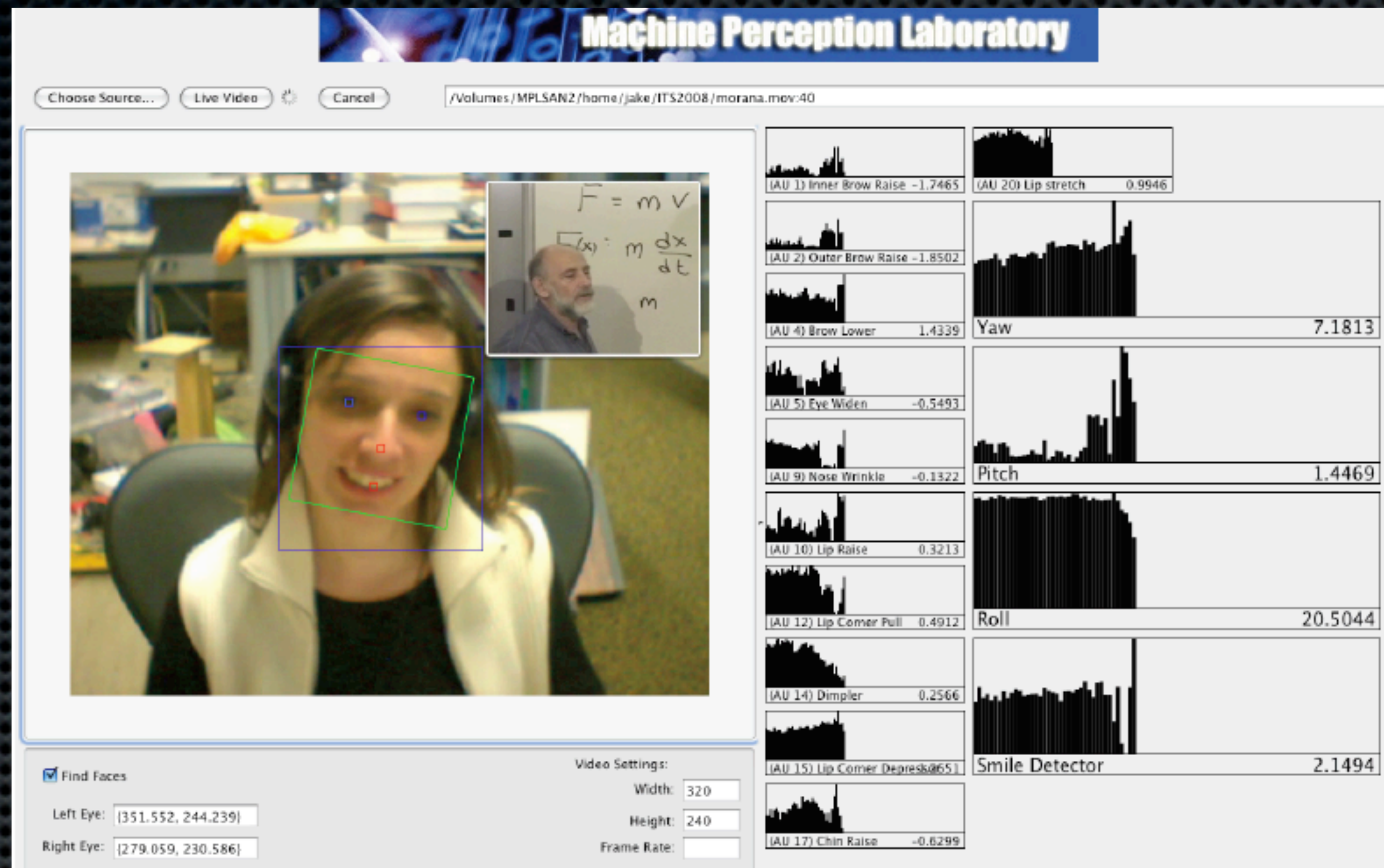
Predicting students' perception of difficulty using automatic facial expression recognition

Perception of topic difficulty

- One important aspect of the student's state is the student's **perception of curriculum difficulty**.
  - A teacher might wish to "slow down" curriculum delivery during parts of a lecture that the student finds difficult.
- We conducted a pilot study in which we used automatic facial expression recognition to predict the student's perception of difficulty.

Pilot study

Whitehill, Bartlett, and Movellan, 2008



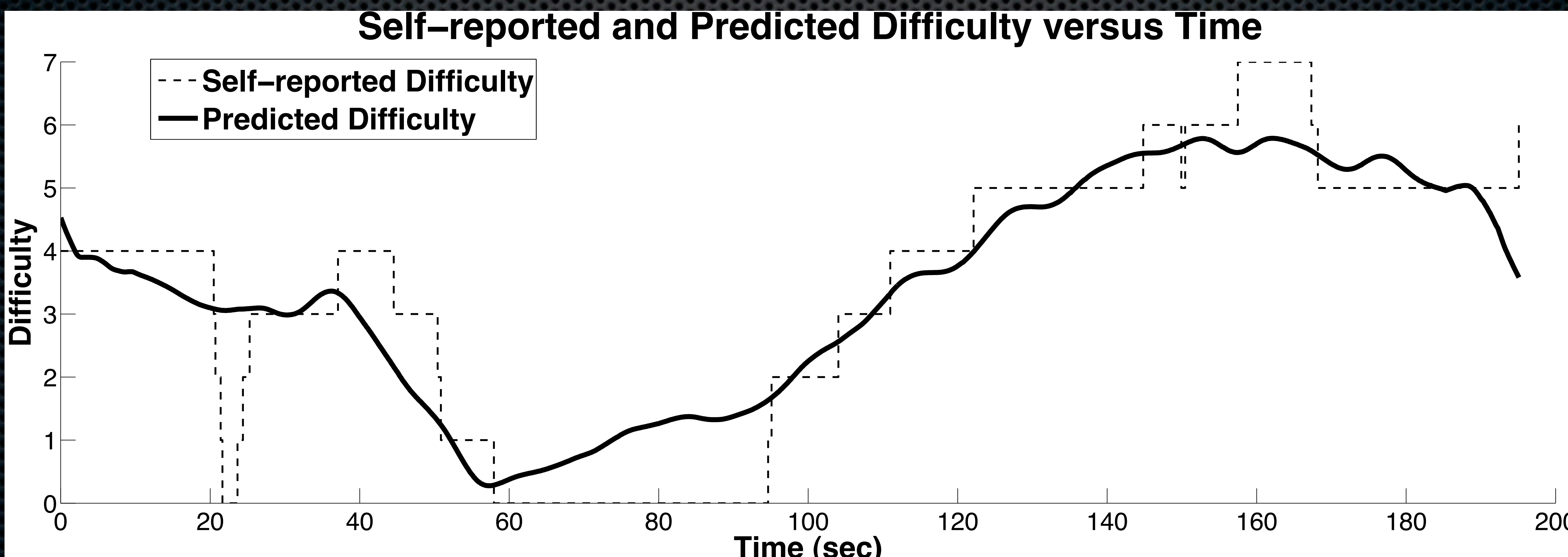
- 8 subjects watched a short video lecture containing segments of mathematics, physics, and psychology.
- Subjects' facial actions were measured automatically by CERT.
- Subjects then watched the lecture video again and rated their perceived difficulty of the lecture at each moment in time.

Pilot study

- For each subject, we trained a regression model to predict the student's perception of topic difficulty on a frame-by-frame basis:
  - **Input features**: CERT's facial action outputs, and their first temporal derivatives.
  - **Target values**: Self-reported perceived difficulty level.
  - **Predictor**: Multivariate linear regression.
- Each subject's data sequence was partitioned into training and testing subsequences.

Pilot study results: prediction accuracy

- Average accuracy on testing sequences (trained on only 100 seconds of video): *R*=0.42 (Pearson), *p*<0.05



Pilot study results: behavior

- The facial actions correlated most highly with perceived difficulty varied substantially from person to person.
- For 6 out of 8 subjects, perceived difficulty was negatively correlated with AU 45 (blink/eye closure).
  - When cognitive load increases, students tend to blink less (Holland and Tarlow, 1972).

Path toward developing an affect-sensitive automated teacher

Proposed path

1. Identify a learning domain in which affect-sensitivity matters.
   - Measure the effect of affect-sensitivity on learning.
2. Record sessions of expert human teachers teaching students.
   - Students' answers to questions, teachers' decisions
   - Video of student
3. Train affective state detectors (e.g., engagement, boredom).
4. Apply machine learning to train an automated affect-sensitive tutor to mimic the pedagogical power of humans.

Investigating the role of affect in cognitive skill training

Cognitive skills training

- Cognitive skills training programs are growing in popularity in middle and high schools.
- Cognitive training games aim to boost students' academic performance by first strengthening basic cognitive processes, e.g.:
  - Working memory
  - Attention
  - Processing speed

Cognitive skills training

- Training of basic cognitive skills is correlated with improved language and reading performance (Temple, et al. 2003) and improved academic performance in underprivileged youth (Turner, Serpell, and Hill 2010).
- Given the benefit of cognitive skills training, it would be useful to automate the process.

Cognitive skills training

- Cognitive skills trainers typically push their students to their limits.
  - Keeping students engaged and motivated may be important factors.
- Cognitive skills training may be a fruitful domain for developing an affect-sensitive teaching system.

Cognitive game example: “Attention Arrows”

www.brainskills.com/shortdemo.html

Cognitive game example: “Attention Arrows”

www.brainskills.com/shortdemo.html

Cognitive training -- human versus computer.

- Hill and Serpell (2009) assessed the effectiveness of **computer-based training** of cognitive skills to **human-based training**.
  - Human tutors delivered higher learning gains.
    - Why?

Hypotheses

- **Performance-sensitivity hypothesis**: Human teachers are very effective at adapting their teaching to the individual skill level of the student.
- **Affect-sensitivity hypothesis**: Human teachers can adapt to the **affective state** of the student.
- **Mere presence hypothesis** (e.g., Guerin 1986): The mere presence of a human observer encourages students to learn better.

Automated cognitive skills training

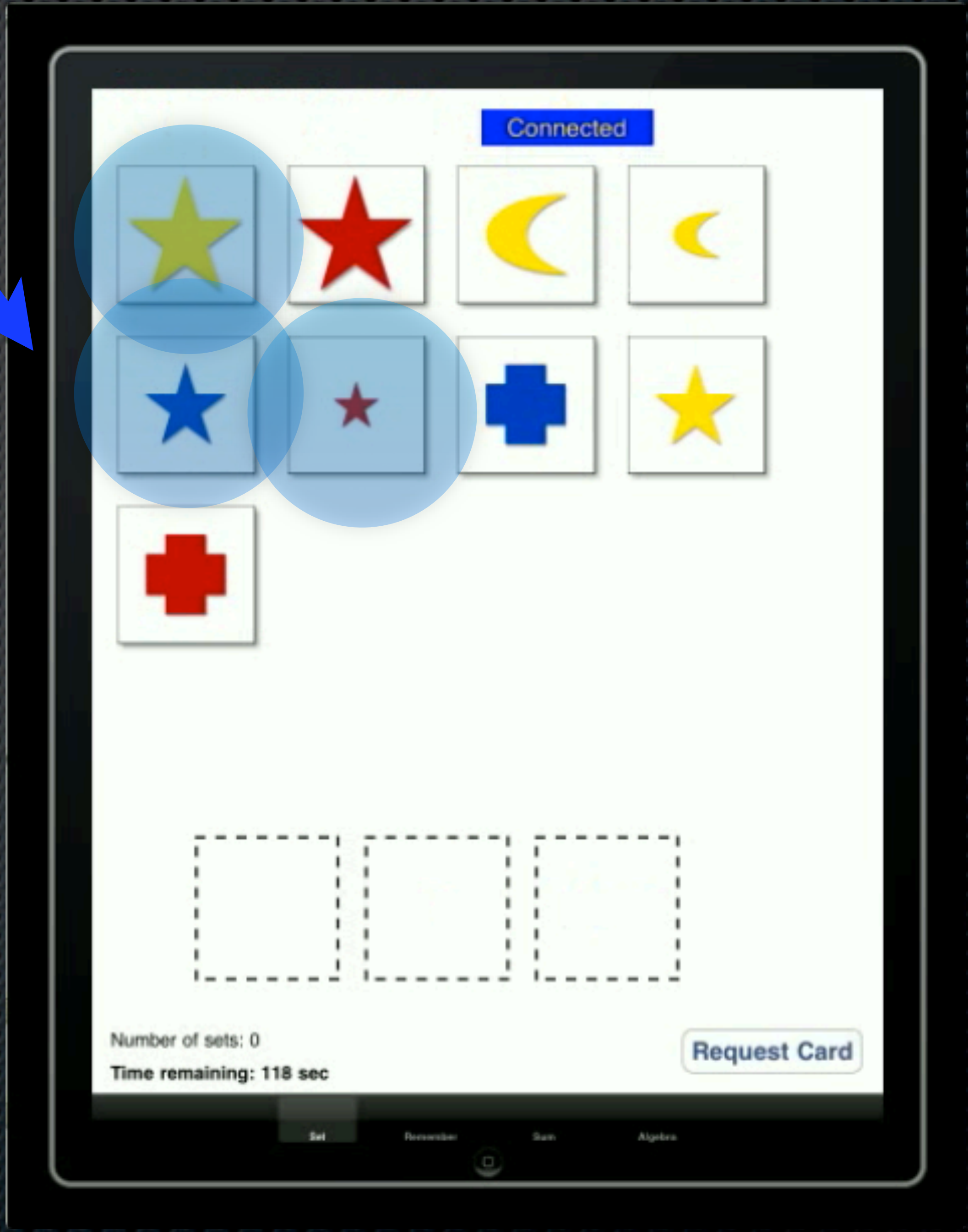
Foster, Lin, Whitehill, Serpell, Bartlett, and Movellan

- In collaboration with the Serpell lab at Virginia State University, we investigated the role of the student's affect in cognitive training.
- We constructed cognitive games training software for the iPad.
  - Collect training data towards developing an automated cognitive skills teacher.

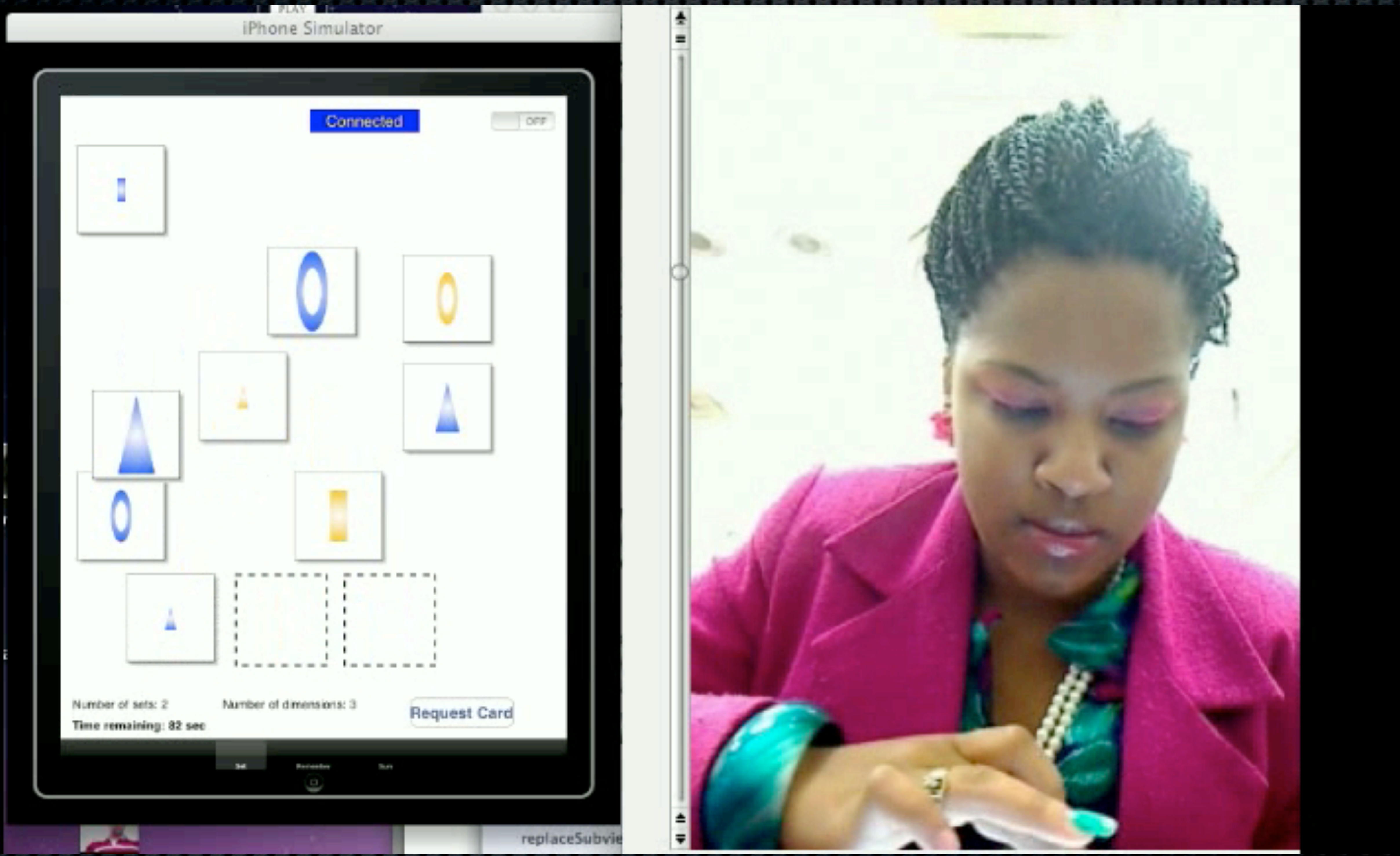
Cognitive games example: “Set”

- Similar to the classic card game.
- Goal: form “sets” of 3 cards each.
- In each set, value of all cards must be all same or all different for **size**, **shape**, and **color**.

Valid set

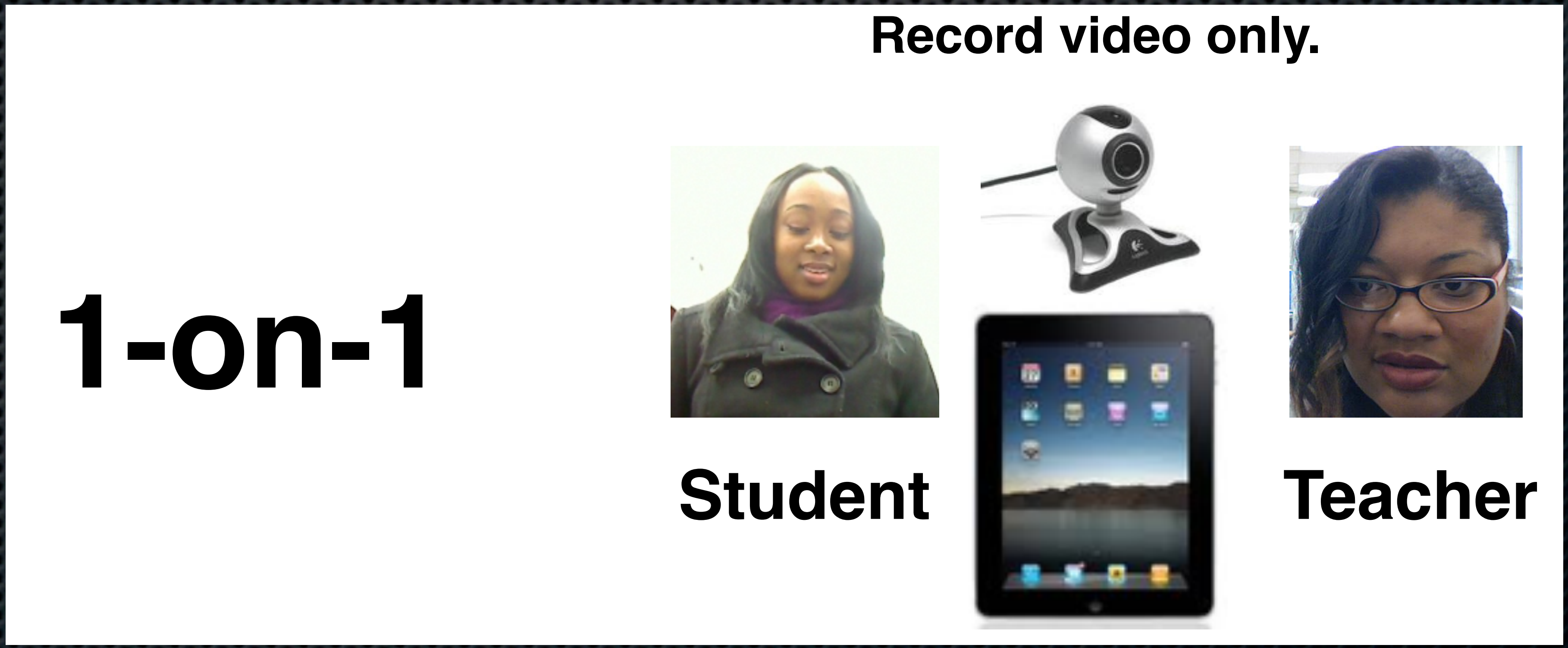


Cognitive training games



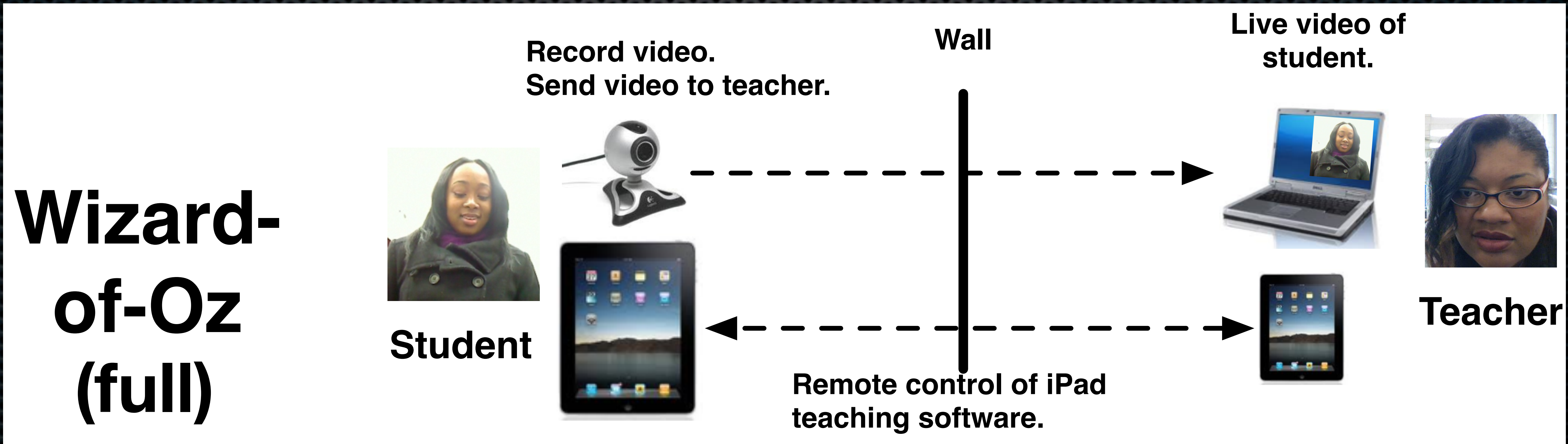
Experimental conditions

- To discern where the learning gains come from, we compared three experimental conditions:
  1. Training 1-on-1 by a human teacher.



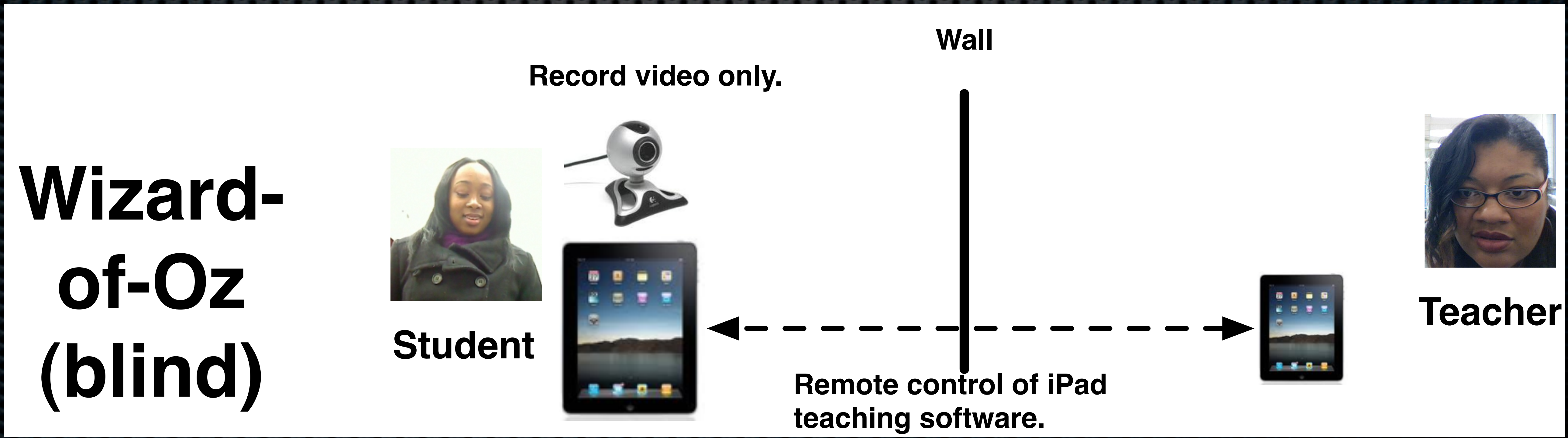
Experimental conditions

2. Training from an “automated” teacher, simulated using a Wizard-of-Oz paradigm.
    - Wizard-of-Oz: the machine is controlled by a human hidden behind a curtain.



Experimental conditions

3. Training from an "automated" teacher but **blind** -- same as #2 above, but teacher could not see the student.



Cognitive games: the teacher's actions

- The teacher's objective is to maximize the student's learning gains only on the **Set** task.
- During the training session, cognitive skills trainers must decide when to:
  - Switch tasks
  - Change task difficulty
  - Give hints
  - Provide encouragement, push student to try harder

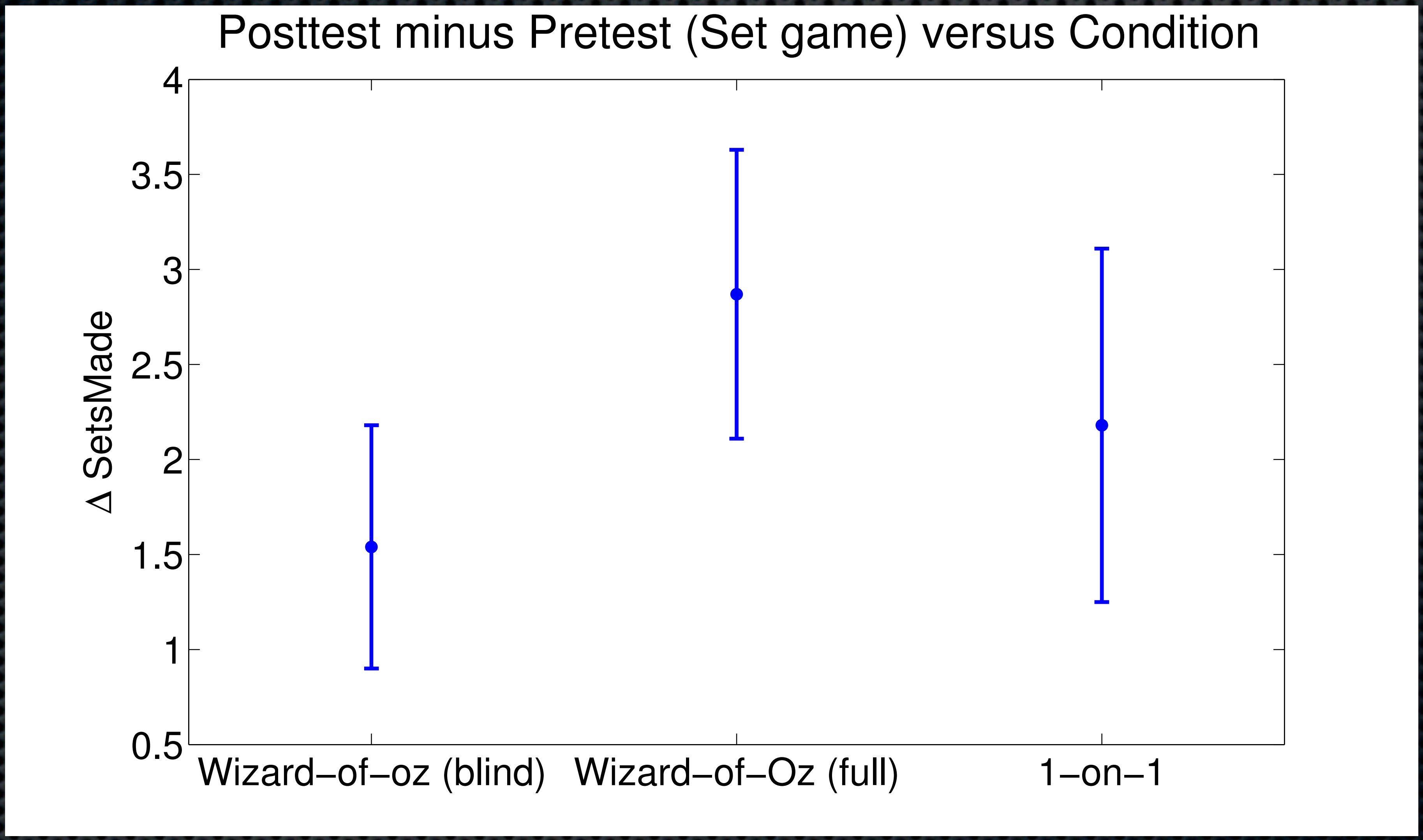
Possible results

- 1-on-1 human > Wizard-of-Oz (full):
  - Supports the **mere presence** hypothesis.
- All three conditions are equal:
  - Supports the **skill level** hypothesis.
- Wizard-of-Oz (full) > Wizard-of-Oz (blind):
  - Supports the **affect-sensitivity** hypothesis.
- 1-on-1 < Wizard-of-Oz conditions:
  - A human's presence could actually decrease learning gains (intimidation?).

Cognitive games experiment

- During Oct-Dec 2010 we conducted a pilot study.
- 66 subjects (51 female, all African-American) participated in cognitive Games training in 1 of the 3 experimental conditions.
- Protocol:
  “Set” **Pretest**.
  ~30 minutes **training** on all cognitive games.
  “Set” **Posttest**.
- We examined differences in learning gains (Posttest minus Pretest) as well as students’ facial expression.

Pilot results: Posttest minus Pretest



Trend: Wizard-of-Oz (full) > Wizard-of-Oz (blind) (p=.09)

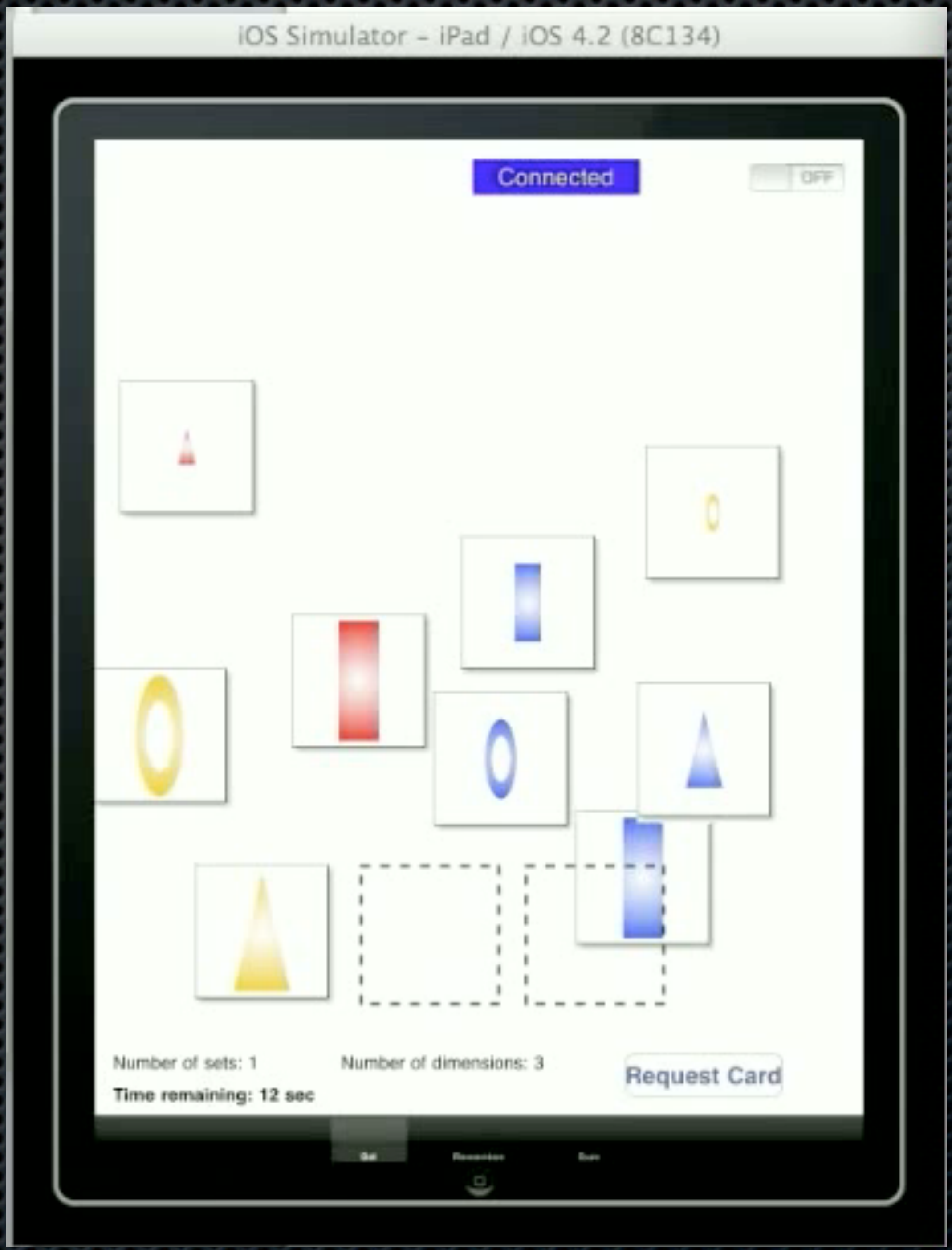
Pattern of results supports the affect-sensitivity hypothesis.

Was there interesting nonverbal behavior?

- How is the student doing?
- Is the student interested?
- Is the task too easy?
- Is the task too hard?
- Is the student trying?

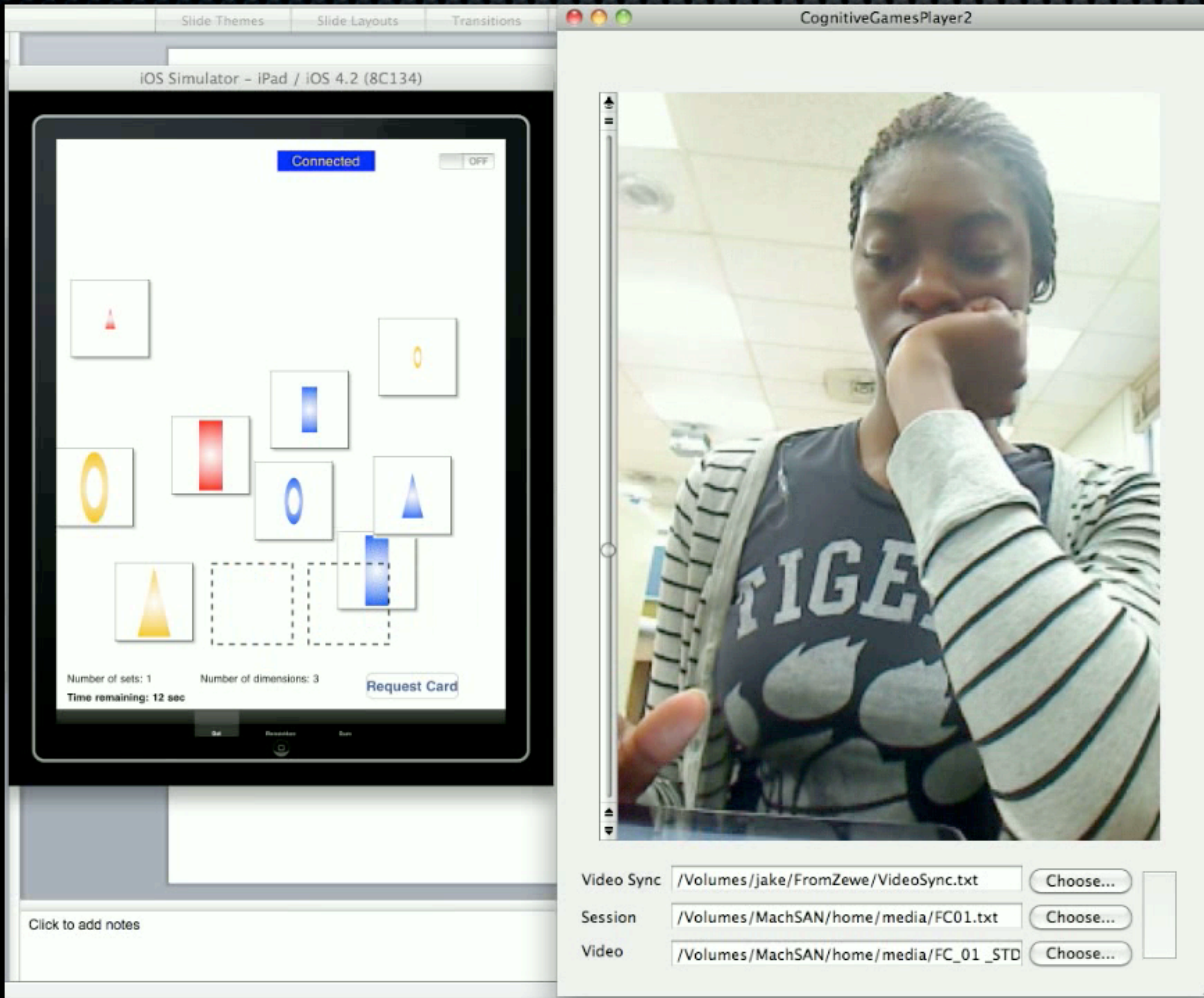
Was there interesting nonverbal behavior?

- How is the student doing?
- Is the student interested?
- Is the task too easy?
- Is the task too hard?
- Is the student trying?

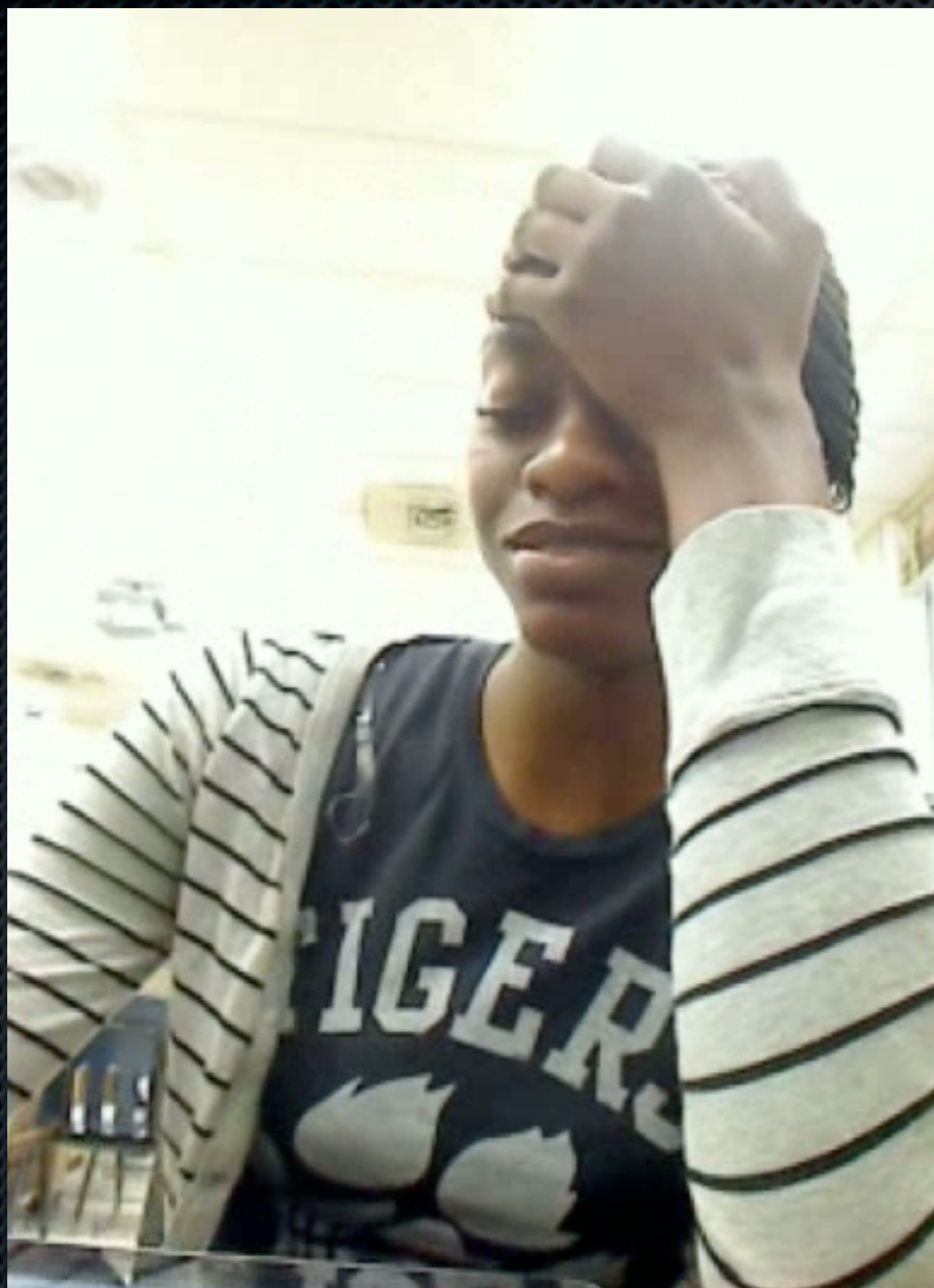


Was there interesting nonverbal behavior?

Was there interesting nonverbal behavior?

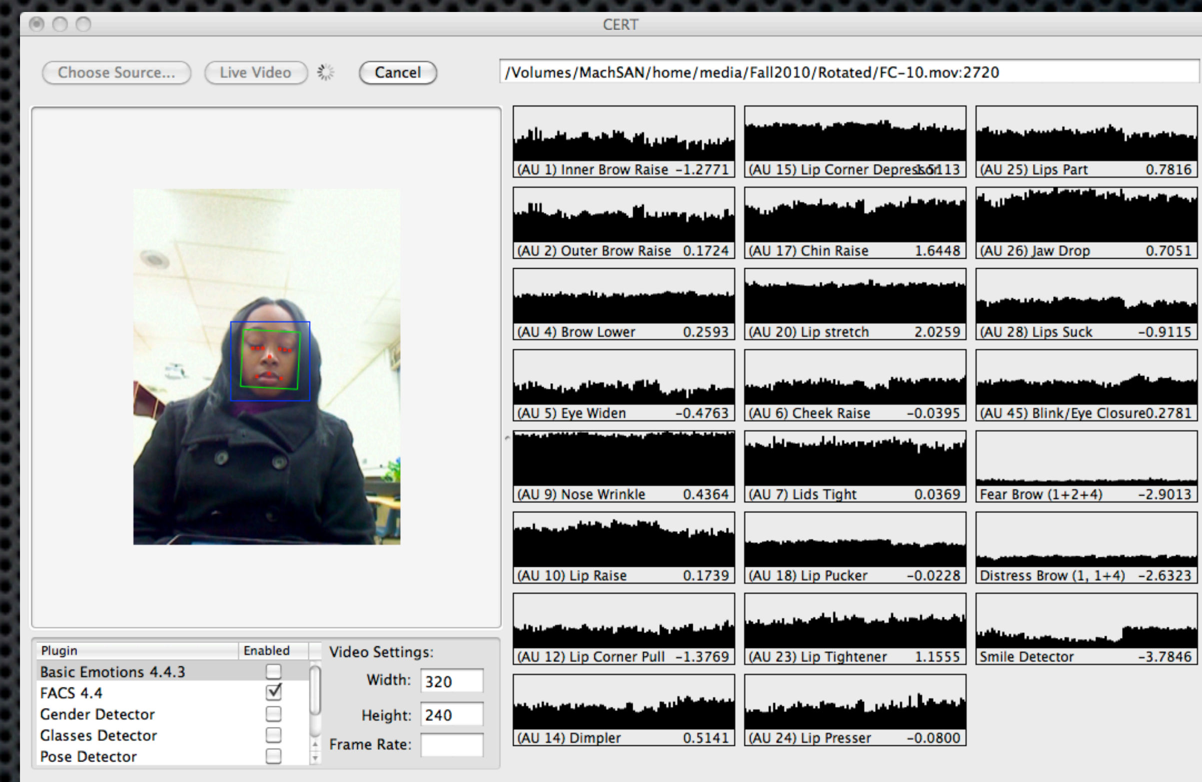


Was there interesting nonverbal behavior?



Automatic analysis of smile

- We studied the correlation of Smile intensity with Posttest minus Pretest performance.
- Smile was coded by the Computer Expression Recognition Toolbox (**CERT**)
    - Fully automatic coding of expression from video (Littlewort, et al., *FG* 2011 Whitehill, et al., *TPAMI* 2009)



Analysis of facial expression: Smile

- Results:
  - Average smile intensity over each training session was negatively correlated (*R*= -0.3353, $p < 0.05$) with Posttest minus Pretest.
    - I.e., subjects who smiled less learned more.
  - Hoque and Picard (FG'2011) found that smile often occurred in natural frustration.

Smile after making an error

Smile after making an error

Engagement

Students who performed well looked engaged and “on-task”, with little variation in affect.

Recognizing student “engagement”

Student “engagement”

- From examining the videos, we found student “engagement” to be an important dimension of variation.
  - Engagement: Is the student immersed in the task and trying to succeed?

Automatic engagement detection

- It may be useful to recognize engagement automatically.
- Automated teaching system:
  - Detect when a student is “gaming” the tutor.
- Human supervision:
  - “Is my kid slouching while doing his math?”

Measuring engagement

- There are two paradigms for measuring student engagement from tutoring sessions:
  - Stimulus-driven measurement
  - Observation-based measurement

Measuring engagement

- There are two paradigms for measuring student engagement from tutoring sessions:
  - Stimulus-driven measurement
  - Observation-based measurement
    - We would be happy for an automated detector to match the perceptual power of a human observer.

Labeling engagement

- Given the observational approach to labeling, how do we code engagement from the videos?
  - How many "levels" of engagement?
    - 3, 4, 5, ..., or infinitely many?
  - How to label:
    - Continuously by turning a "dial" while watching video
    - Watch video segment, then give single number
  - Timescale of labeling:
    - 60 sec video segment? 10 sec?

Labeling engagement

- Based on inter-coder reliability and labelers' feedback, we chose:
  - 4 discrete engagement levels.
  - Give a single number after watching a video chunk.

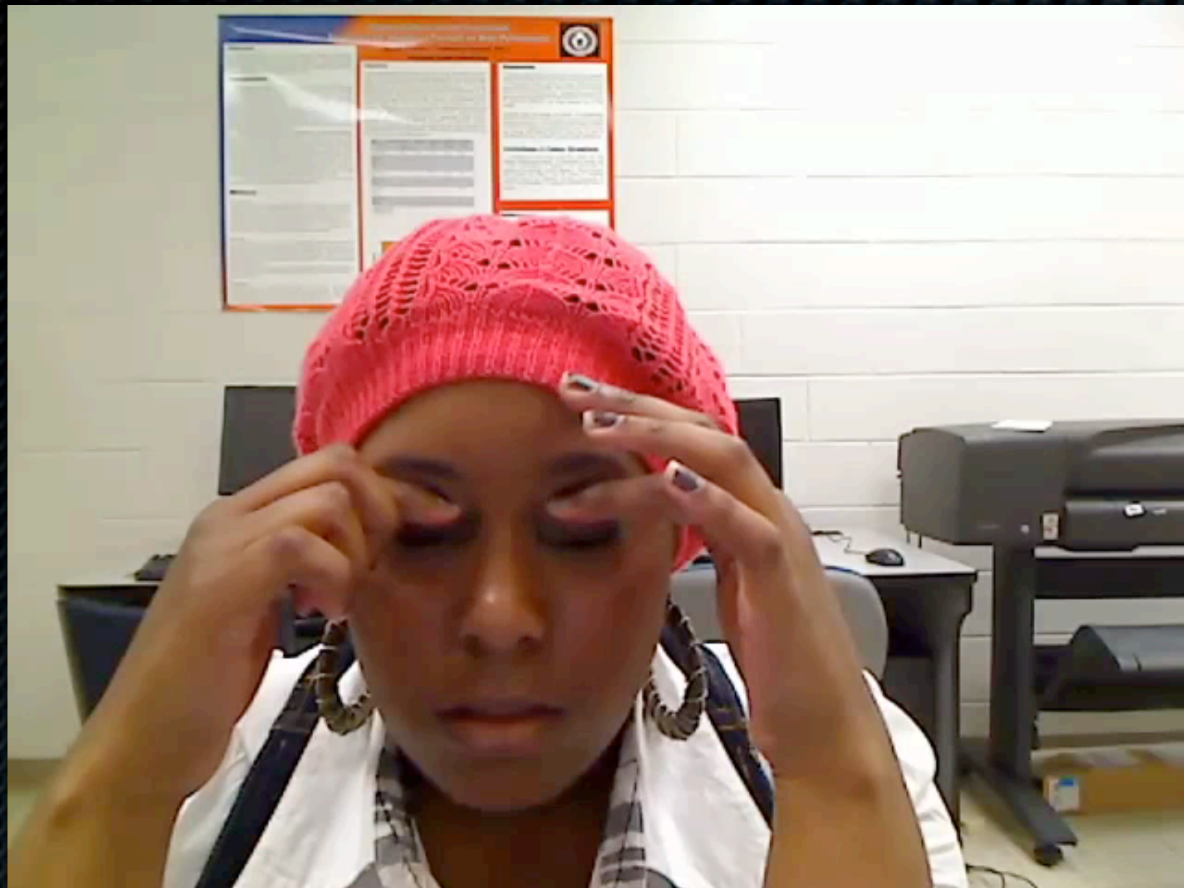
Labeling engagement

- Engagement labels guidelines:
  - 1 = Not engaged at all, e.g., looking away from computer and obviously not thinking about task, eyes completely closed.
  - 2 = Nominally engaged, e.g., eyes barely open, clearly not “into” the task.
  - 3 = Engaged in task -- student requires no admonition to “stay on task”
  - 4 = Very engaged -- student could be “commended” for their level of engagement in task.

Labeling engagement

- We compared reliability of labeling 10sec versus 60sec chunks of video for “engagement”.
  - 60sec: kappa = 0.39
    - Problem: What if student was engaged early on and disengaged toward the end?
  - 10sec: kappa = 0.72

1



2

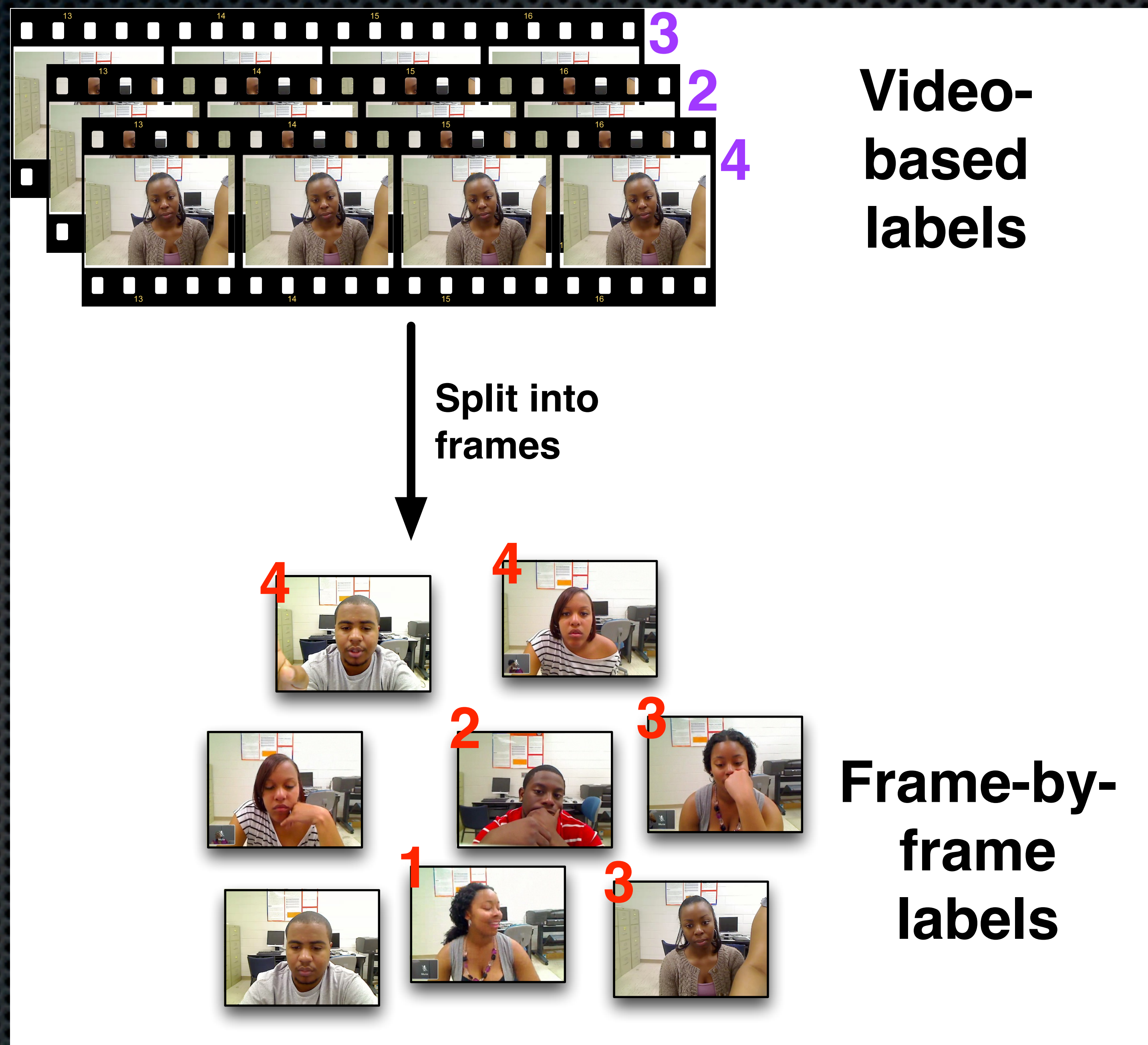
3

4

Frame-by-frame labeling

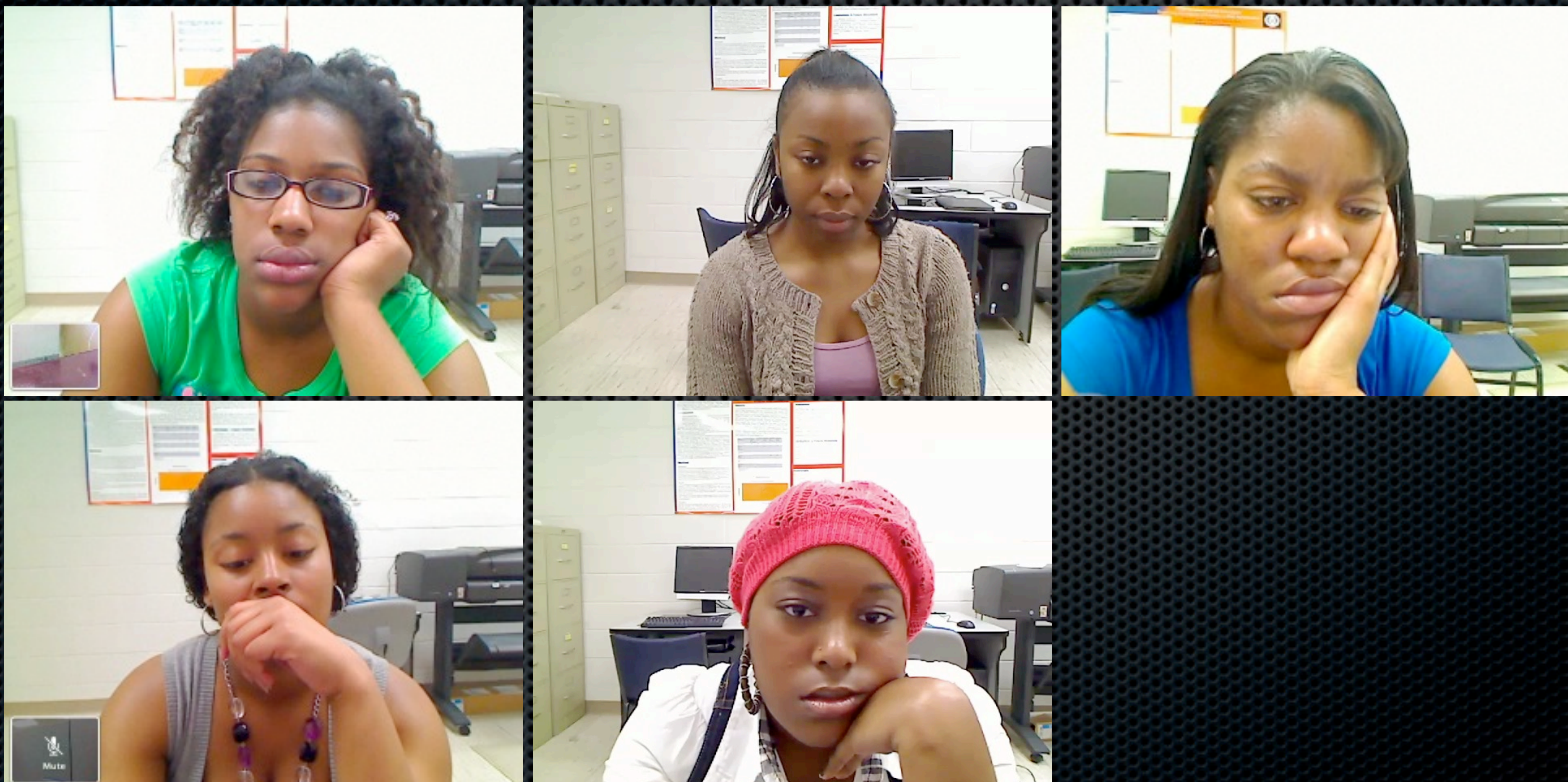
- What about even smaller timescales?
- The shortest possible timescale is a single video frame, captured every 1/30th of a second.
- How reliably do frame-by-frame labels of engagement predict the labels given to a 10sec video clip?
  - How much engagement information is contained in time dynamics compared to static appearance?
  - A frame-based automatic engagement detector may be easier to train.

Frame-by-frame labeling



Engagement level 1

Engagement level 2



Engagement level 3

Engagement level 4

Frame-by-frame labeling

- We split 120 video clips (each 10sec long, from 25 subjects of cognitive training videos) into 40 frames each.
- 4800 video frames were randomly shuffled over time and subject.
- Coders labeled randomized frames in batches of 100.
- Coders' video labels were then reconstructed by averaging their frame levels for each video.
    - Kappa = 0.75

Frame-by-frame labeling

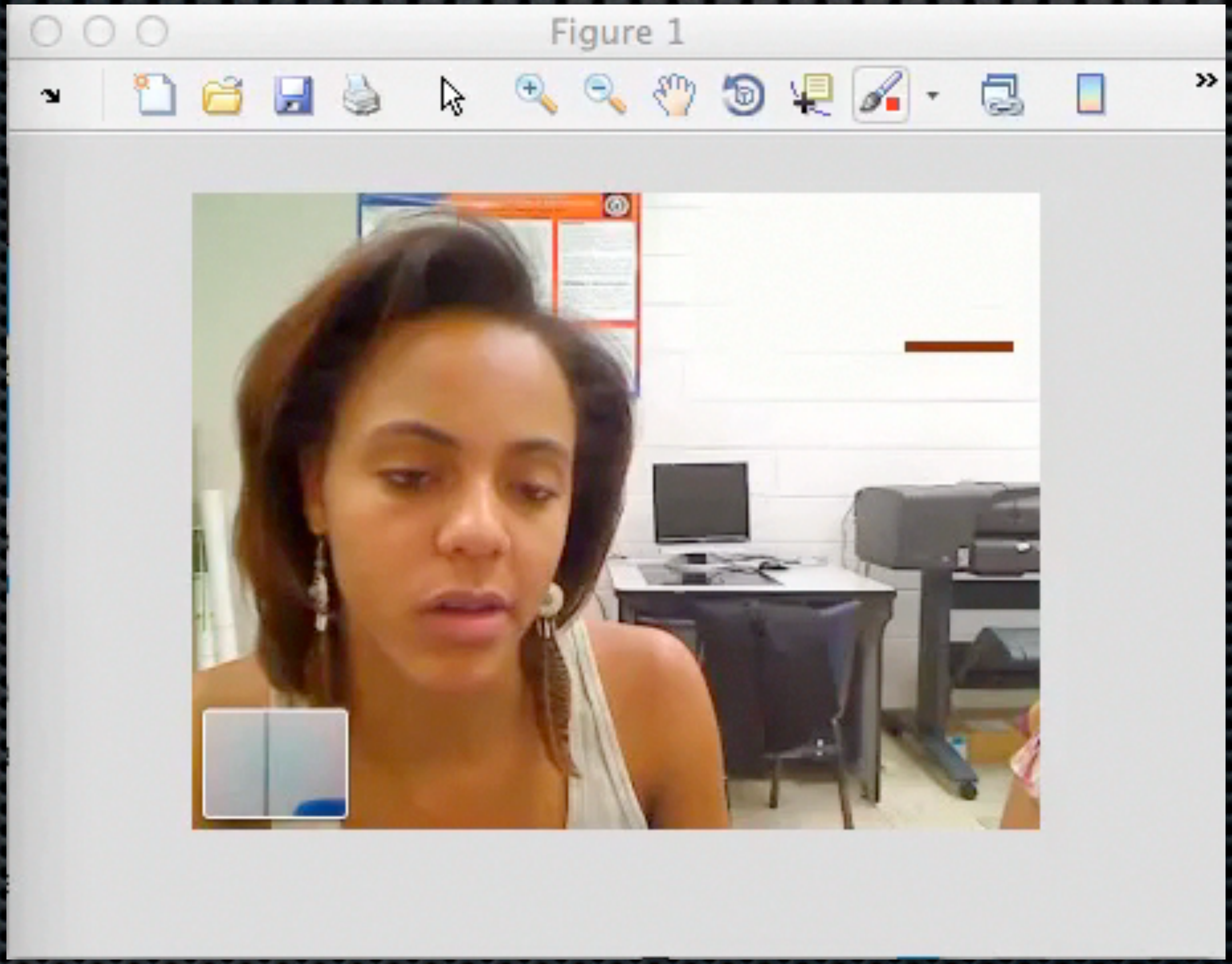
- Reliability of frame-by-frame labeling suggests that most of the time dynamics information about engagement may be correlated with static, appearance-based information.
- In this learning environment, frame-based detection of engagement may work well.

Automatic engagement detection

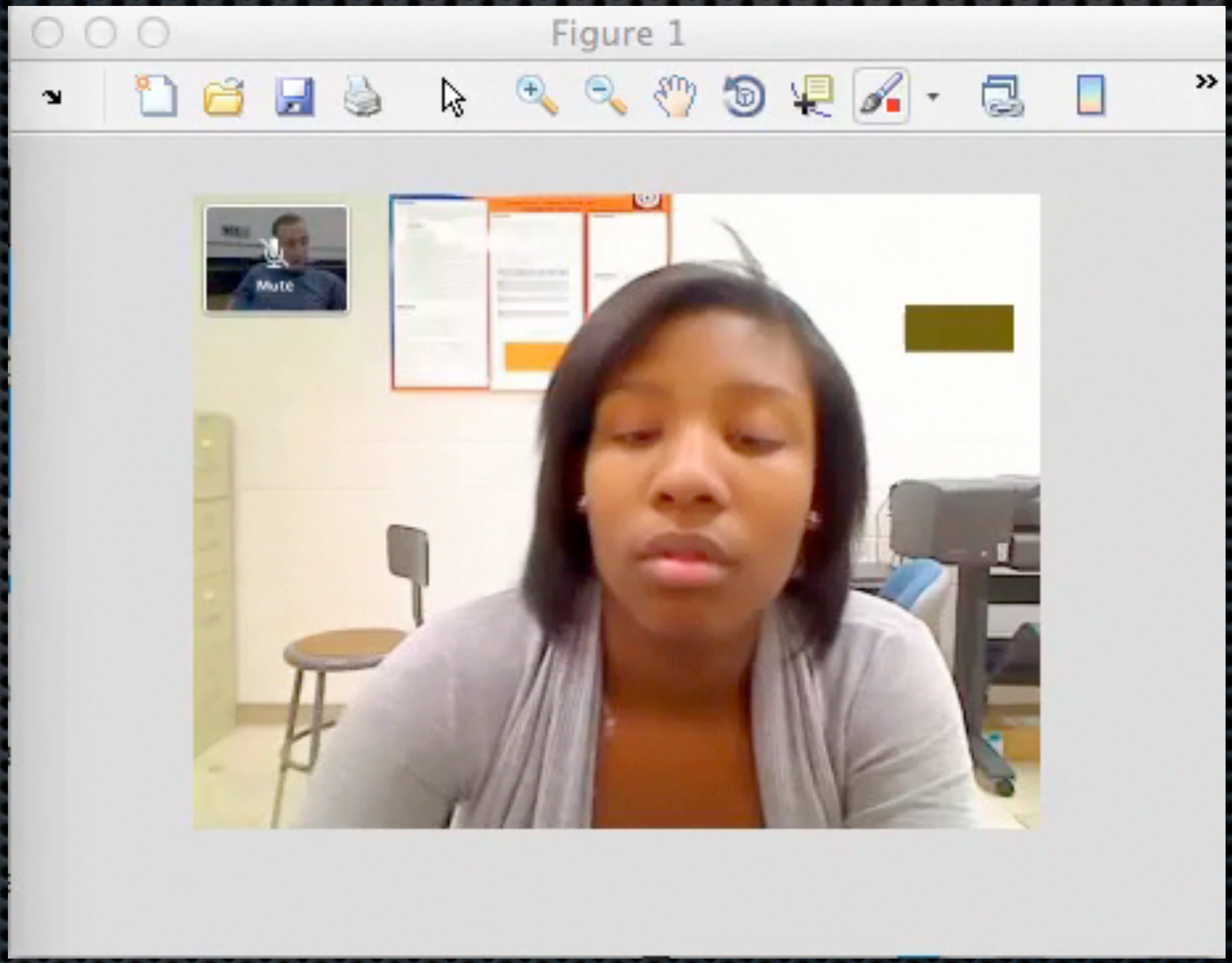
- We developed automatic, frame-by-frame detectors of engagement:
  - In each frame, find the face
  - From the face, estimate facial action units (AUs)
  - Using multivariate logistic regression and AU estimates, estimate engagement level of each frame.

Automatic engagement detection

Automatic engagement detection



Automatic engagement detection



Engagement analysis

- Using machine learning and CERT, we can analyze *how* humans made their judgments about level of engagement.
  - Which action units were correlated with student engagement?
  - For distinguishing engagement=1 from {2,3,4}, most predictive features:
    - Eye blink
    - Pose (roll) of the head

Ongoing and future work

Synopsis

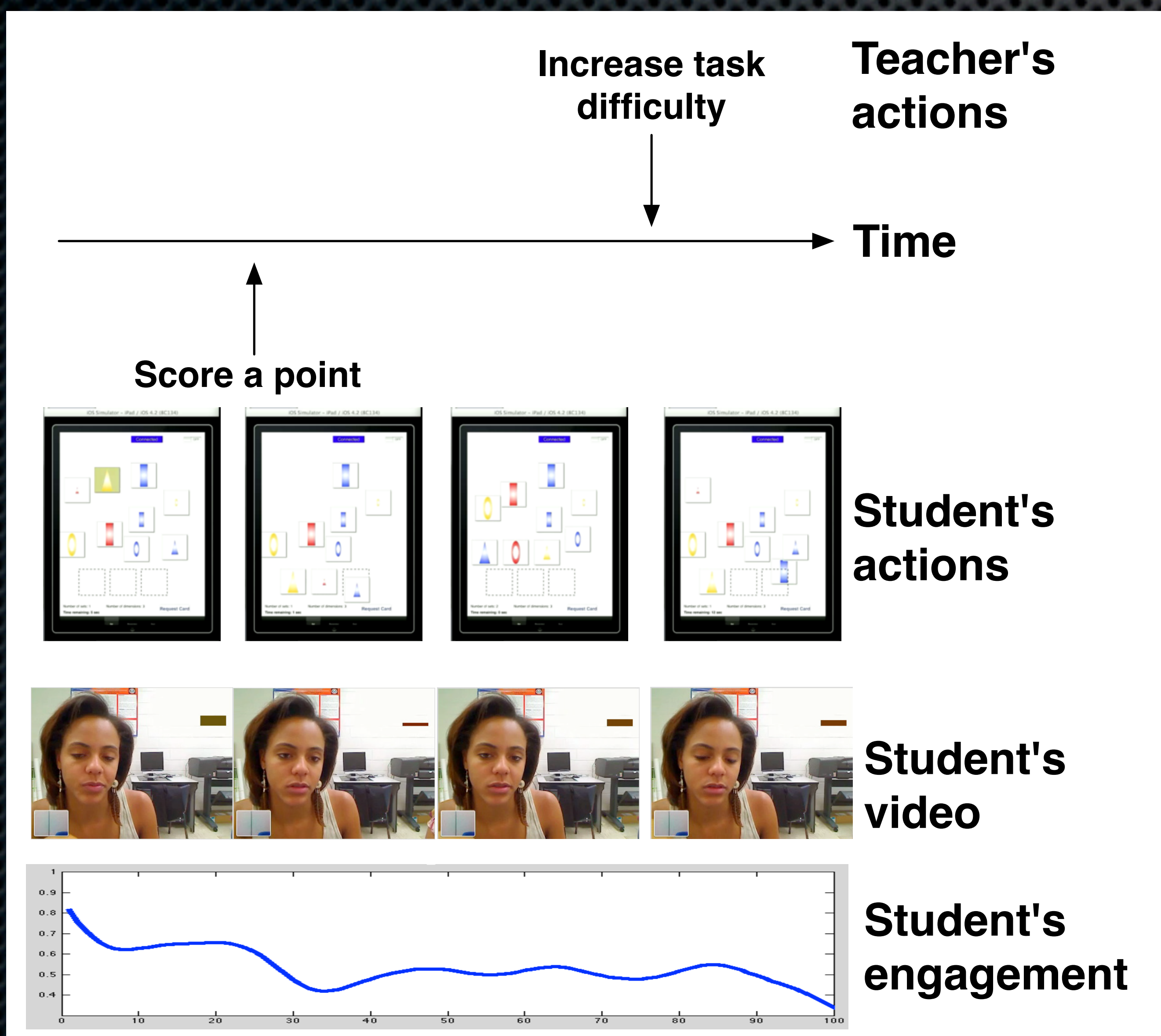
- We have identified a learning domain -- cognitive skills training -- for which there is evidence that affect-sensitivity makes a difference.
- On the other hand: differences in learning gains not as strong as we had hoped.
- We were glad that subjects were at least learning!
- Interesting affect:
  - “Catastrophic” episodes
  - “TSA problem”

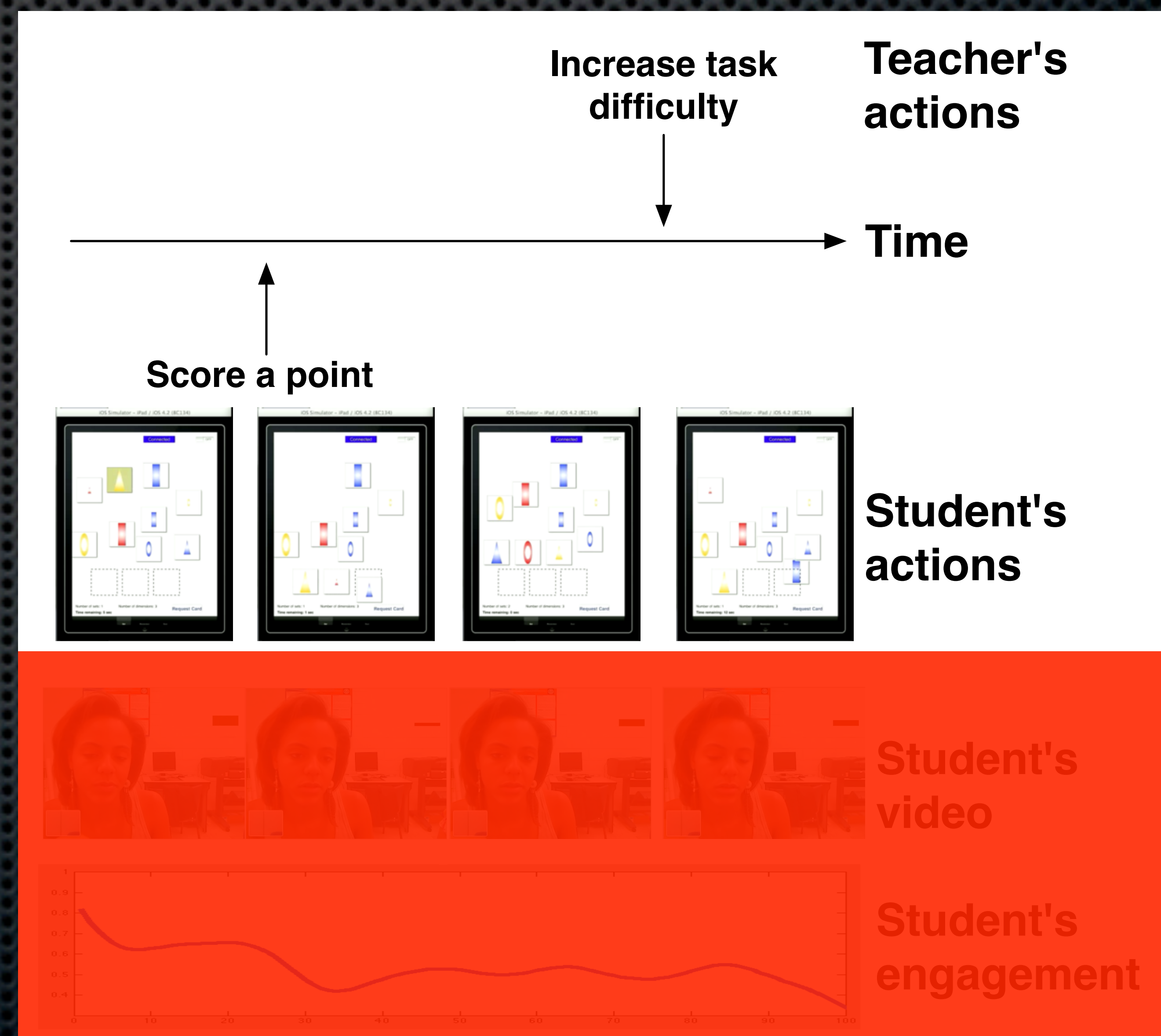
Teaching analysis

- Instead of focusing on learning gains, we can also examine:
  - Does affect-sensitivity affect *how* teachers teach?
  - Do teachers' control policies vary by condition?
- Using machine learning, we can train a model to predict the teacher's next action given the history of student's and teacher's prior actions.
  - If a model trained from affect-sensitive teachers fit the affect-blind teaching sessions poorly, then that suggests that affect makes a difference.

Teaching analysis

- As another measure of the effect of affect on teaching, we can train a model to predict the teacher's next action using two alternative different sets of features:





Teaching synthesis

- Ultimately, we want to synthesize teacher actions in a way similar to how expert human teachers teach.
  - E.g., we can train a classifier to predict when to give a hint given the history of student & teacher actions.
  - Suppose a classifier has an area-under-ROC of 0.9:
    - Is that good?
    - How to evaluate?
    - Ultimately we care about learning, not accuracy in reproducing the trainer's actions.

END