Proximate Measures of Engagement, Learning, and Relationships for High-Impact Tutoring

Researcher Breakout
NSSA Conference

May 9, 2023

NATIONAL STUDENT SUPPORT ACCELERATOR

equalizing access to quality tutoring

Guiding questions

1) What are the proximate measures of (a) engagement, (b) relationships, and (c) learning that you find compelling?

2) What are the questions that are most exciting to us that these measures enable?

3) How can we share better share measures with one another?

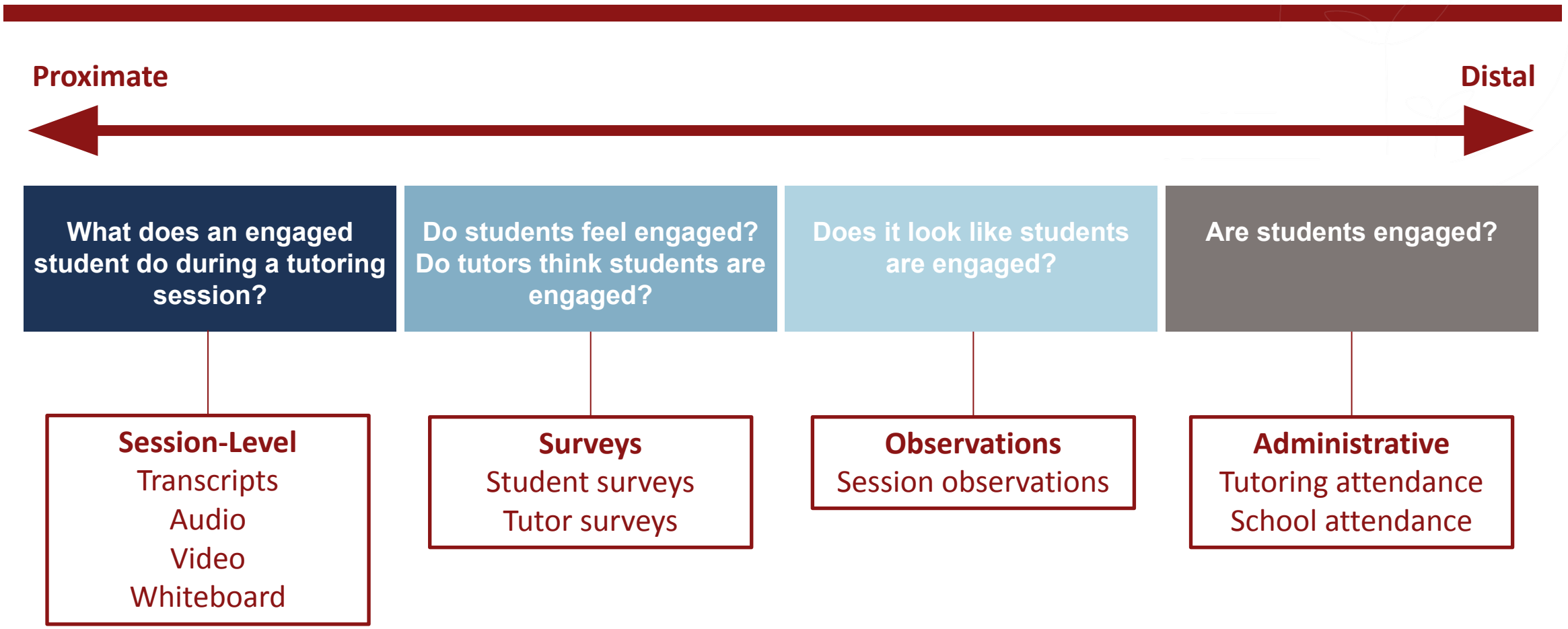
- Open-source materials
- Facilitate interpreting and comparing results

Status quo: How does tutoring affect student outcomes?

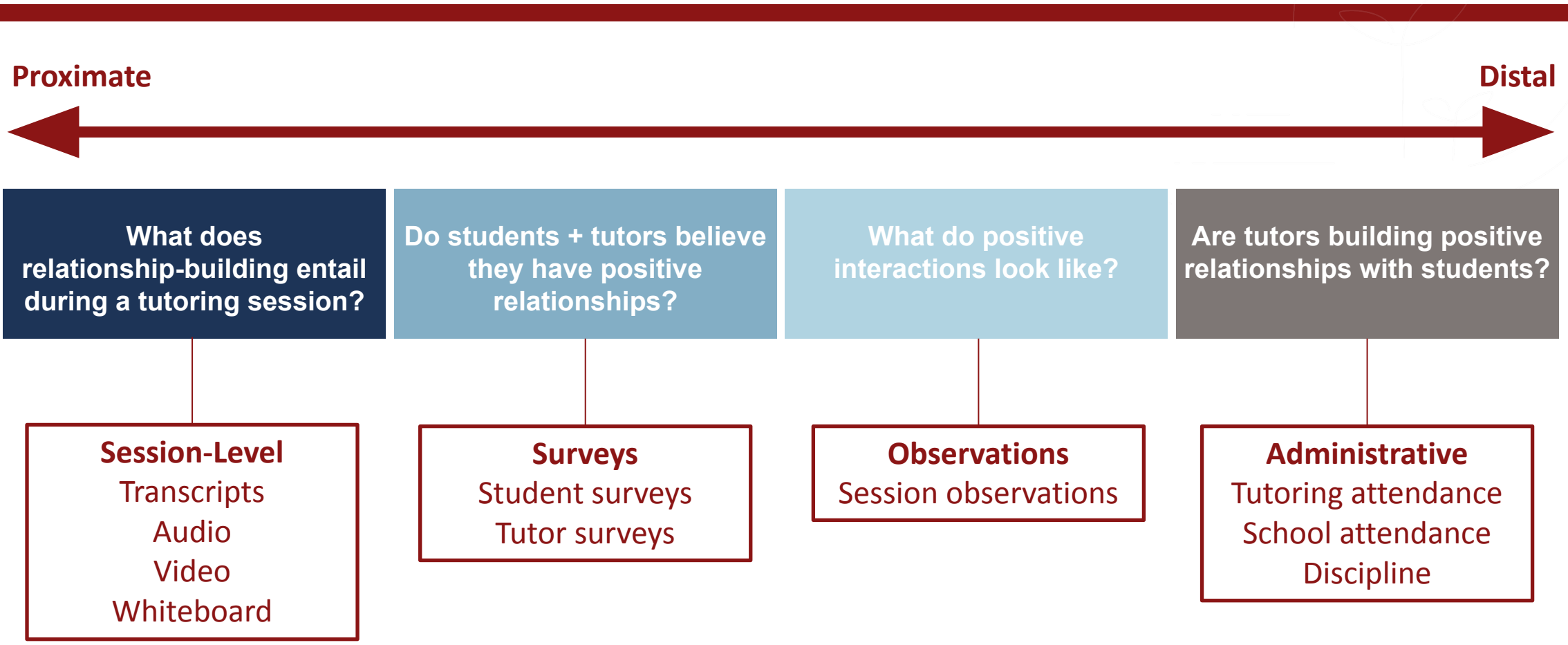
- Engagement
  - Tutoring attendance
  - School attendance
- Relationships
  - Tutoring attendance
  - School attendance
  - Discipline
- Learning
  - Grades
  - End-of-year assessments

Can we learn what is going on faster and with more fine-grained data?

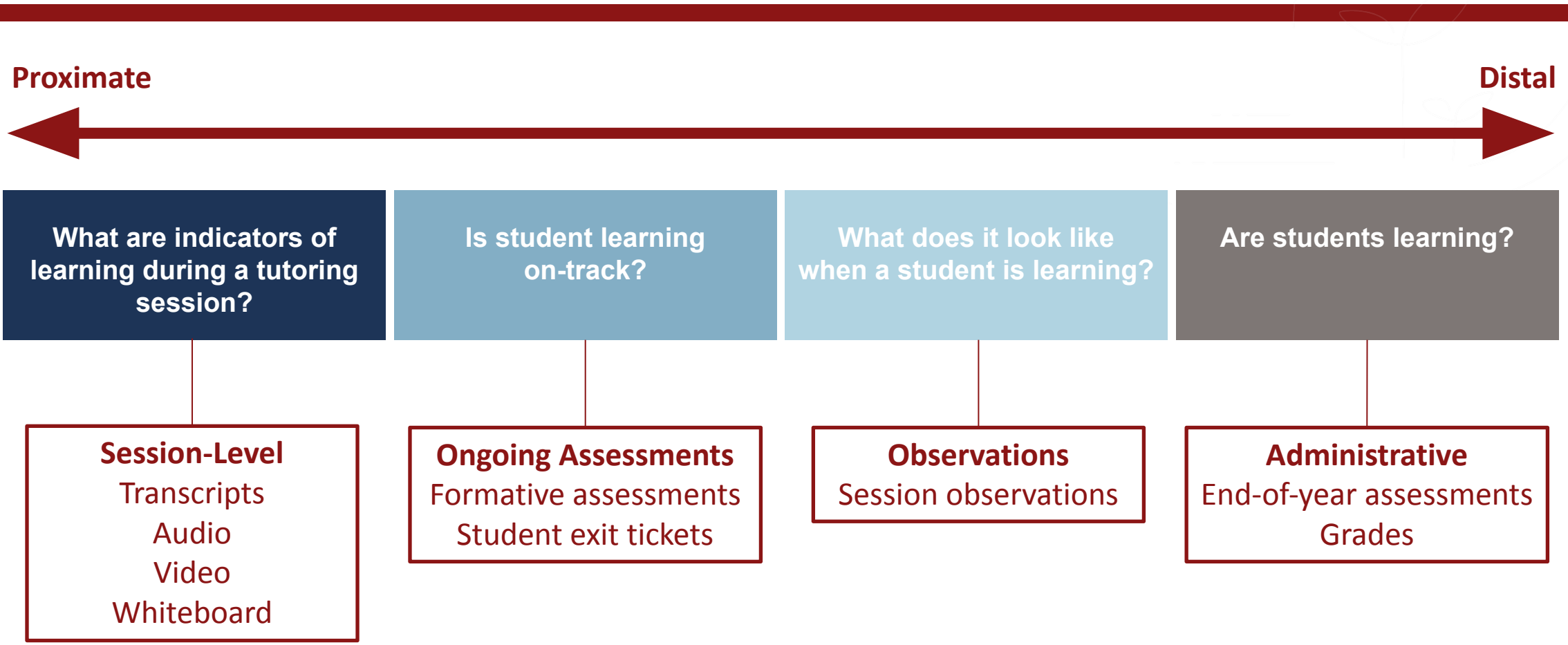
Measuring engagement



Measuring relationships



Measuring learning



Observations, Surveys, Ongoing Assessments

What does an engaged student do during a tutoring session?	Do students feel engaged? Do tutors think students are engaged?	Does it look like students are engaged?	Are students engaged?
What does relationship-building entail during a tutoring session?	Do students + tutors believe they have positive relationships?	What do positive interactions look like?	Are tutors building positive relationships with students?
What are indicators of learning during a tutoring session?	Is student learning on-track?	What does it look like when a student is learning?	Are students learning?
Session-Level Transcripts Audio Video Whiteboard	**Surveys** Student surveys Tutor surveys **Ongoing Assessments** Formative assessments Student exit tickets	**Observations** Session observations	**Administrative** End-of-year assessments Grades School Attendance Tutoring Attendance Discipline

Observations

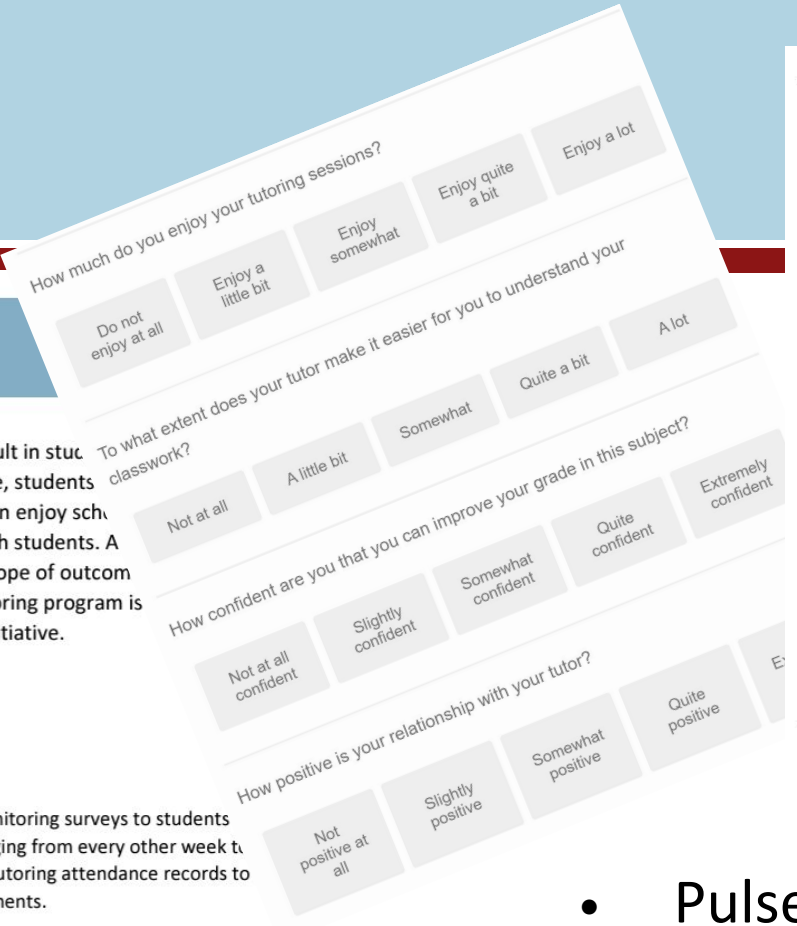
- Observations have largely been used to measure teacher effectiveness
- Pros:
  - Quantifies educator behaviors beyond administrative data
  - Measures interactions with students
- Cons:
  - Infrequent
  - Most are measuring the educator/instruction, not the student
  - No one really is looking at whether individual kids are paying attention
  - Hard to scale
  - Subjective

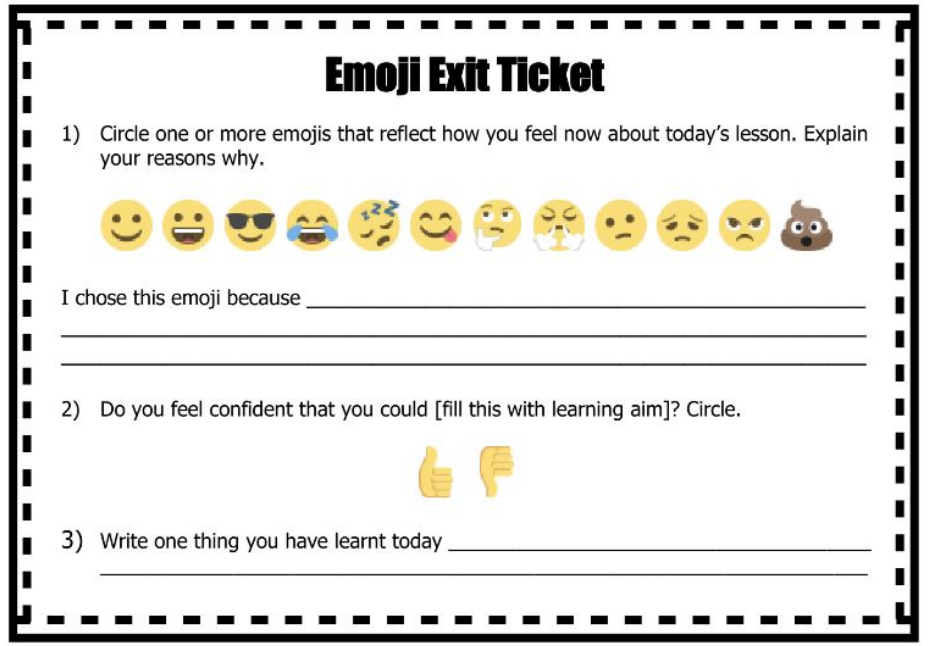
Surveys

- Students and tutors respond to items that give insights into their perceptions and beliefs
- Pros:
  - People – even students! – will tell you how they are feeling
  - People's perceptions interact with their behaviors
- Cons:
  - Social desirability bias
  - Asking the right questions
  - Survey fatigue/Often low participation
  - Surveys can be interventions themselves

Surveys

Tutoring Surveys





Emoji Exit Ticket

1) Circle one or more emojis that reflect how you feel now about today's lesson. Explain your reasons why.

I chose this emoji because _____

2) Do you feel confident that you could [fill this with learning aim]? Circle.

3) Write one thing you have learnt today _____

While research shows that high-impact tutoring programs consistently result in stuc the benefits of tutoring can extend beyond academic learning. For instance, students more value in the subject, develop a connection with a caring adult, or even enjoy sch also stand to benefit from their experiences instructing and interacting with students. A student and tutor surveys provides an opportunity to capture a broader scope of outcom scores or grades. Moreover, surveys can start to shed light as to why a tutoring program is successful and provide action-oriented insights into how to improve the initiative.

Types of surveys (and when to administer them)

Pulse Checks

To gauge how a tutoring program is going, you can administer brief, progress monitoring surveys to students and tutors. These pulse checks should be administered on a consistent basis, ranging from every other week t every month. Program leaders can pair the data from the frequent surveys with tutoring attendance records to take stock of how the tutoring program is proceeding and make real-time adjustments.

Extended Surveys

The extended surveys aim to assess a much broader range of outcomes. These surveys contain more questions and take longer to complete, so they should be used more sparingly. In general, it is recommended to administer the extended student and tutor surveys at the beginning and end of a tutoring program to assess changes over time. Importantly, the questions in these surveys can be administered to all students, not just those who receive tutoring (unless otherwise noted).

How to administer surveys

While you can always use paper and pen to administer surveys, the easiest way to collect data is to use an online survey platform. Some popular online survey platform options with free account options include Google Forms, Survey Monkey, and Qualtrics.[1]

[1] These items were created to assess the target constructs using survey design best practices. Many of the scales were adapted from Panorama Education's Social-Emotional Learning Survey (www.panoramaed.com).

- Pulse checks
  - Post-session
  - Several times a year
- Extended surveys
  - Mid-year; End-of-year

Example: Tips4Tutors intervention focused on SEL

Relationships
(Weeks 1, 7, 9, 11)

Tutor-student relationships

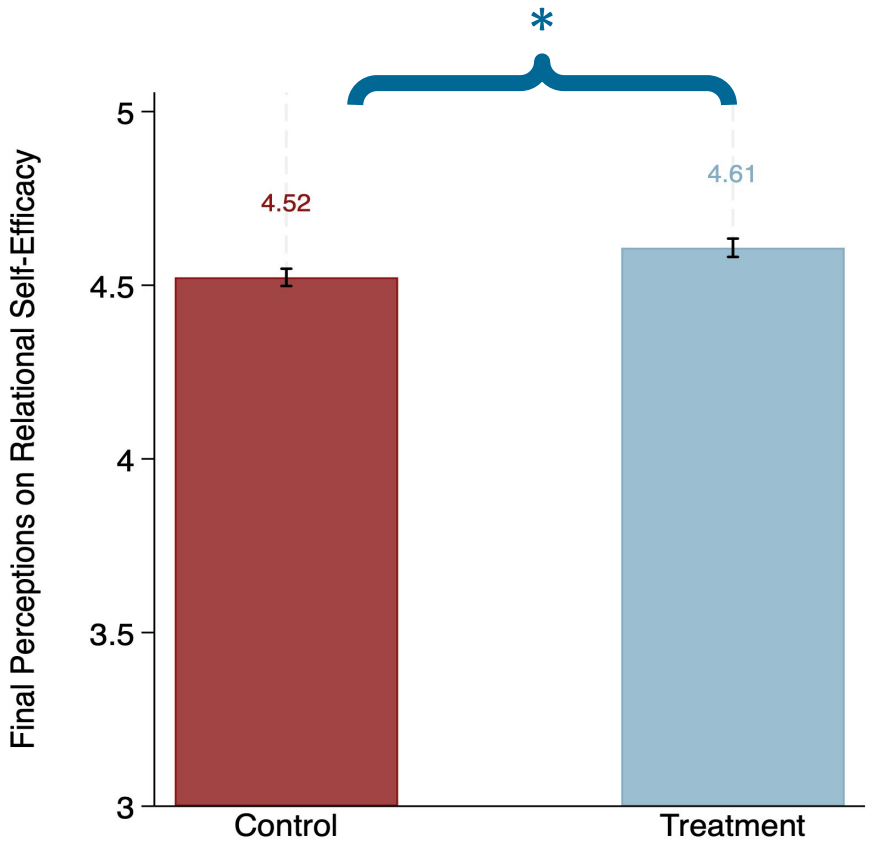
Relational self-efficacy

WARM UP:
ASK Qs TO LEARN MORE

Ask students one of these questions, or think of your own:

- What superpower would you want to have and why?
- What is one thing you like to do outside of school?
- What is one thing you like about school?
- When you are feeling down, what makes you feel better?

How confident are you that you can build positive relationships with the students you tutor?



ES = 0.18 sd

Example: Tips4Tutors intervention focused on SEL

Stress Management
(Weeks 4, 5, 6, 10)

Student stress management

Tutor stress management

WARM UP:
STRESS CODEWORD

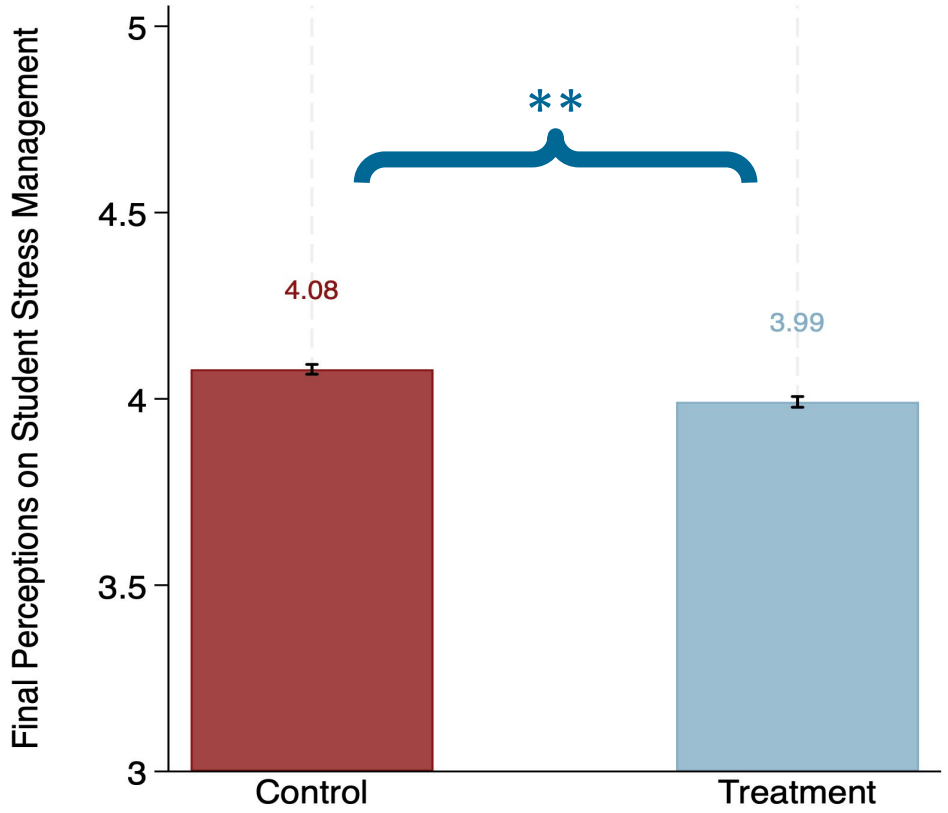
Come up with a word or signal to relax stress. Get creative!

Pause! Boink Banana Freeze

Students can use the codeword whenever they need a break during a tutoring session.

Encourage them to try it out during your session today!

How confident are you that you can support students to develop coping strategies?



ES = -0.12 sd

Exit Tickets

- Get session-level feedback on whether students are understanding concepts
- Pros:
  - Provides real-time data on student mastery of concepts
- Cons:
  - Time constraints
  - Hard to compare across students

What about session-level data & novel methods…

What does an engaged student do during a tutoring session?	Do students feel engaged? Do tutors think students are engaged?	Does it look like students are engaged?	Are students engaged?
What does relationship-building entail during a tutoring session?	Do students + tutors believe they have positive relationships?	What do positive interactions look like?	Are tutors building positive relationships with students?
What are indicators of learning during a tutoring session?	Is student learning on-track?	What does it look like when a student is learning?	Are students learning?
Session-Level Transcripts Audio Video Whiteboard	**Surveys** Student surveys Tutor surveys **Ongoing Assessments** Formative assessments Student exit tickets	**Observations** Session observations	**Administrative** End-of-year assessments Grades School Attendance Tutoring Attendance Discipline

Using Natural Language Processing to Measure Engagement, Relationships and Learning

Dora Demszky, Assistant Professor, Stanford Graduate School of Education

NSSA Conference, May 9 2024

Improving lives through learning

Stanford | GRADUATE SCHOOL OF EDUCATION

What's the richest source of naturally occurring data that can reveal **why and how** students are engaging and learning?

Stanford | GRADUATE SCHOOL OF EDUCATION

What's the richest source of naturally occurring data that can reveal **why and how** students are engaging and learning?

Tutor-student discourse!

NLP can help us **scalably** examine tutor student discourse

Construct	Example Measures
Engagement	Duration and sequencing of turns; Addressee detection; Student questioning and reasoning
Relationships	Emotion detection; Building common ground; Linguistic accommodation; Topic dynamics
Learning	Student reasoning;

NLP can help us **scalably** examine tutor student discourse

Construct	Example Measures
Engagement	Duration and sequencing of turns; Addressee detection; Student questioning and reasoning
Relationships	Emotion detection; Building common ground; Linguistic accommodation; Topic
Learning	Student reasoning;

Unsupervised methods (no need to train own measure)

NLP can help us **scalably** examine tutor student discourse

Construct	Example Measures
Engagement	Duration and sequencing of turns; Addressee detection; Student questioning and reasoning
Relationships	Emotion detection; Building common ground; Linguistic accommodation; Topic
	Training classifiers
Learning	Student reasoning;

Stanford | GRADUATE SCHOOL OF EDUCATION

Fine-tuning LLMs for **measurement**

1. Obtain **representative data**
2. Work closely with experts to develop a **coding protocol** for the construct
   a. often highly subjective
   b. several cycles of training → annotation → discussion
3. **Annotate at least a few hundred examples for each category** within a construct
   a. your total data will be a few hundred examples x number of categories (e.g. yes/no, not helpful/neutral/helpful etc.)
4. **Fine-tune** pre-trained language model
   a. in our experience, RoBERTa tends to work best, better than GPT-4
5. Rigorously **evaluate** model performance
   a. compare to experts
   b. validate against other data (ratings of instruction

Stanford | GRADUATE SCHOOL OF EDUCATION

We have created measures for…

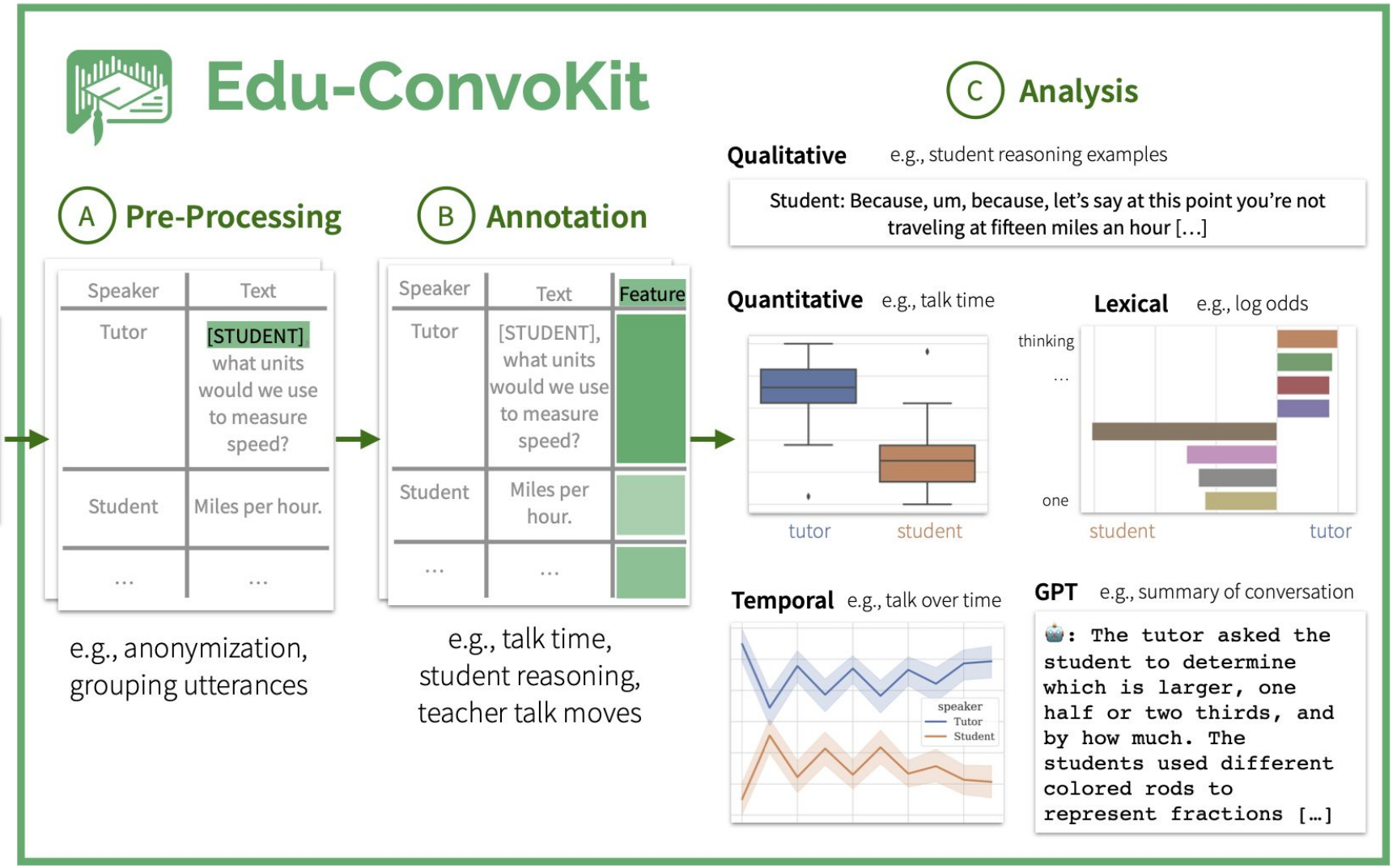
- Uptake of student ideas
- Attributing ideas to students
- Funneling vs focusing questions
- Student reasoning
- On vs off task instruction
- Punitive vs restorative behavior management
- Accountable talk moves (CS instruction)
  - Eliciting student ideas
  - Building on student ideas
  - Connecting ideas

Learn more in our papers!

Stanford | GRADUATE SCHOOL OF EDUCATION

Try out EduConvokit!

Rose Wang



Stanford | GRADUATE SCHOOL OF EDUCATION

Questions **for You**

1. What **language-based measures** of (a) engagement (b) relationships and (c) learning do you find compelling and why?
2. What type of **challenges** have you encountered in creating or using such measures?
3. How can we **share** measures so that we can compare results?

Audio & Visual

Jing Liu

National Student Support Accelerator

equalizing access to quality tutoring

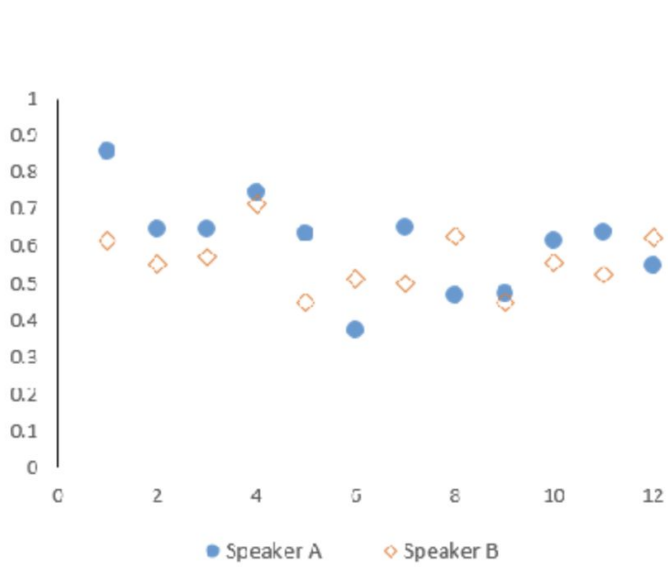
Additional Data Sources: Audio & Video

- Audio and video recordings contain rich information of tutor-student interactions that is not available in other data sources
- Virtual tutoring already collects such data automatically, but researcher just begin to tap into such data and understand their uses in educational settings
- Hypothesis: such data might be particularly useful when language data are sparse or limited (e.g., early literacy tutoring transcript)

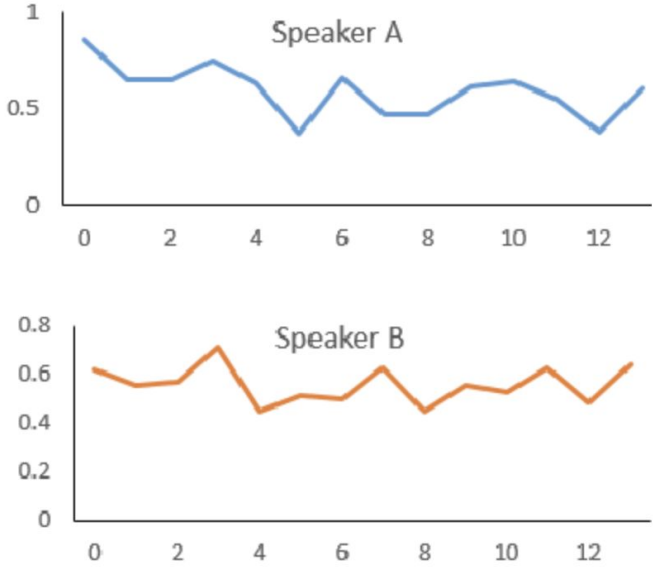
S3 20:43 Groans, "Sweet are not on the list." "Can we get jelly?"
S1 21:00 Begs.
S3 21:01 Begs.
S1 21:04 Star.
S3 21:05 Star.
S1 21:16 Dad.
S3 21:21 Dad.
S1 21:25 Agrees.
S3 21:26 Agrees. "Yes, jelly--"
S1 21:33 Jam.
S3 21:34 "Jam is on the list. [We will?]." Star--
S1 22:02 Gets a fright. There is.
S3 22:18 A.
S1 22:21 Bright.
S3 22:24 Bright green.
S1 22:29 Dragon in the jar.
S3 22:39 In the jar.
S1 22:46 The.
S3 22:48 The.
S1 22:49 Dragon swoops.
S3 22:54 Swoops.
S1 22:56 Down.
S3 22:56 Down.
S1 22:58 And chomps a bun. It glugs.
S3 23:07 It glugs.
S1 23:09 Some sweet.
S3 23:10 Some sweet.
S1 23:11 Milk.
S2 23:13 Milk.
S1 23:16 Star.
S3 23:19 Trails [crosstalk].
S1 23:22 Trails the dragon. Looks at the trail of food. Very good. Okay. Where are Dad and Star? Where are they?
S3 23:42 At the store.
S1 23:43 At the store, very good. And why does Dad say no to crisps?
S3 23:50 Because it's sweet?
S1 23:52 That's good. They're not on the list, are they?
S3 23:57 No.
S1 23:59 Yeah, they're not on the list. So what gives Star a fright?
S3 24:05 It's [inaudible].

Audio Data to Detect Rapport Building

- "In-sync" with each other: when people become more similar to each other in their pitch, loudness, or speaking rate during the course of a conversation (acoustic-prosodic entrainment), there is more likely to be rapport reported by speaker partners (Lubold et al, 2014)
- Can we look at how tutors build rapport with students?



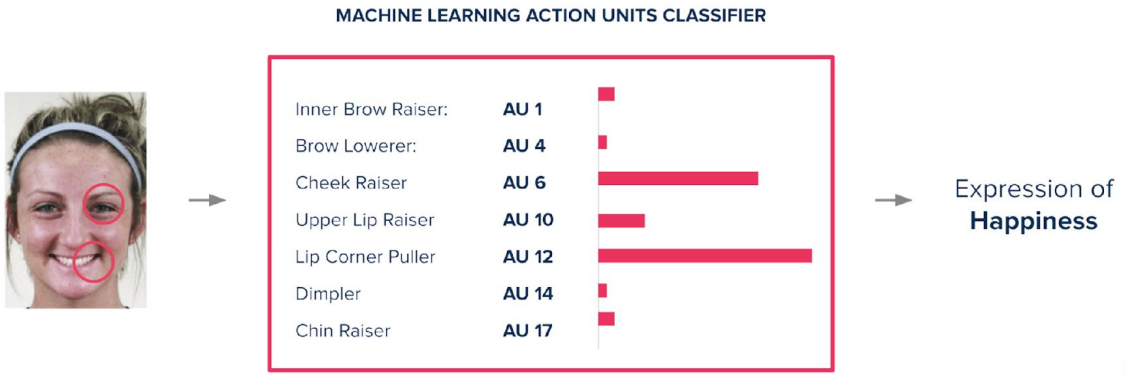
(a) Proximity & Convergence



(b) Synchrony

Video Data for Emotion Detection

- Researchers typical measure facial action units, the movement of a facial muscle or muscle groups that configure the expression of an emotion.
- Most such applications are for adults, not children.
- Can we measure student emotions and provide feedback to tutors in real time?



Considerations When Using Audio and Video Data

- How might audio and video data complement other types of data, especially when measuring similar constructs (e.g., emotions)?
- How can we build domain-specific measures for teaching and learning (e.g., engagement)?
- What are the potential uses of measures derived from audio and video, especially if we can do it in real time?
- Privacy issues: both types of data are PII
- Data storage and computation requirement are very high

Guiding Questions

NATIONAL STUDENT SUPPORT ACCELERATOR

equalizing access to quality tutoring

Guiding questions

1) What are the proximate measures of (a) engagement, (b) relationships, and (c) learning that you find compelling?

2) What are the questions that are most exciting to us that these measures enable?

3) How can we share better share measures with one another?

- Open-source materials
- Facilitate interpreting and comparing results