

INVITED KEYNOTE

Process Data Before and After AI

Continuity · Change · Comparability



Okan Bulut

Professor of Measurement, Evaluation, and Data Science
University of Alberta

“

**Much of what we know about process data
was learned in a world that **no longer exists.****

A starting premise





Direction of process data research in three terms



Continuity

What carries forward: The analytic toolkit largely survives.



Change

What is genuinely new: Process data becomes linguistic.



Comparability

What we can no longer assume: Findings may not transfer.

The promise: *My talk ends in an agenda, not an alarm.*

PART I

What we've learned so far: the foundation





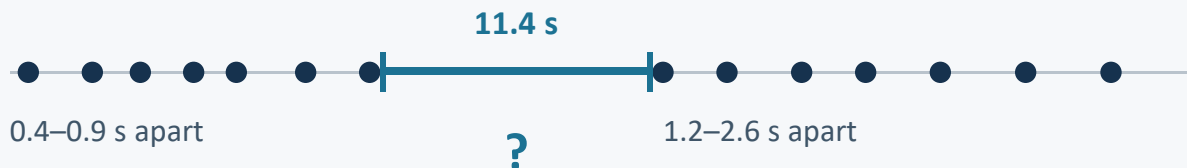
We learned to detect, not to interpret

A decade of process data mined the **exhaust of the task**: response times and clickstreams.

We became good at seeing **that** something happened, and less good at saying **why**.

RESULT: Purely empirical, methodological studies with minimal to no use of learning theories.

One learner's session, as the log records it



The pause is logged. The reason is not.

DETECT

rapid guessing · disengagement · recurring action sequences

INTERPRET

meaning, intent, mechanism



A decade of process data, mapped

232 studies, six recurring research topics [*\(Anghel, Khorramdel & von Davier, 2024\)*](#)



Theory is the exception. Only in digital writing do most studies build on one; elsewhere the field describes behaviour more than it explains it.



Timing & disengagement

- Slow responses surfaced through sequential process analysis in international large-scale assessments
- Foundational lessons from response-time and clickstream data for better psychometrics
- Speed matters in both directions...
 - **Too fast** → rapid guessing
 - **Too slow** → insufficient effort, (productive) struggling, daydreaming, so on...

	Too Fast	Optimal Speed	Too Slow
Correct			
Incorrect			

[Bulut \(2022, ETS Process Data SIG\)](#) · [Jerez, Mazzullo & Bulut \(2026\)](#)

What timing buys

Engagement signals

Rapid guessing, non-response

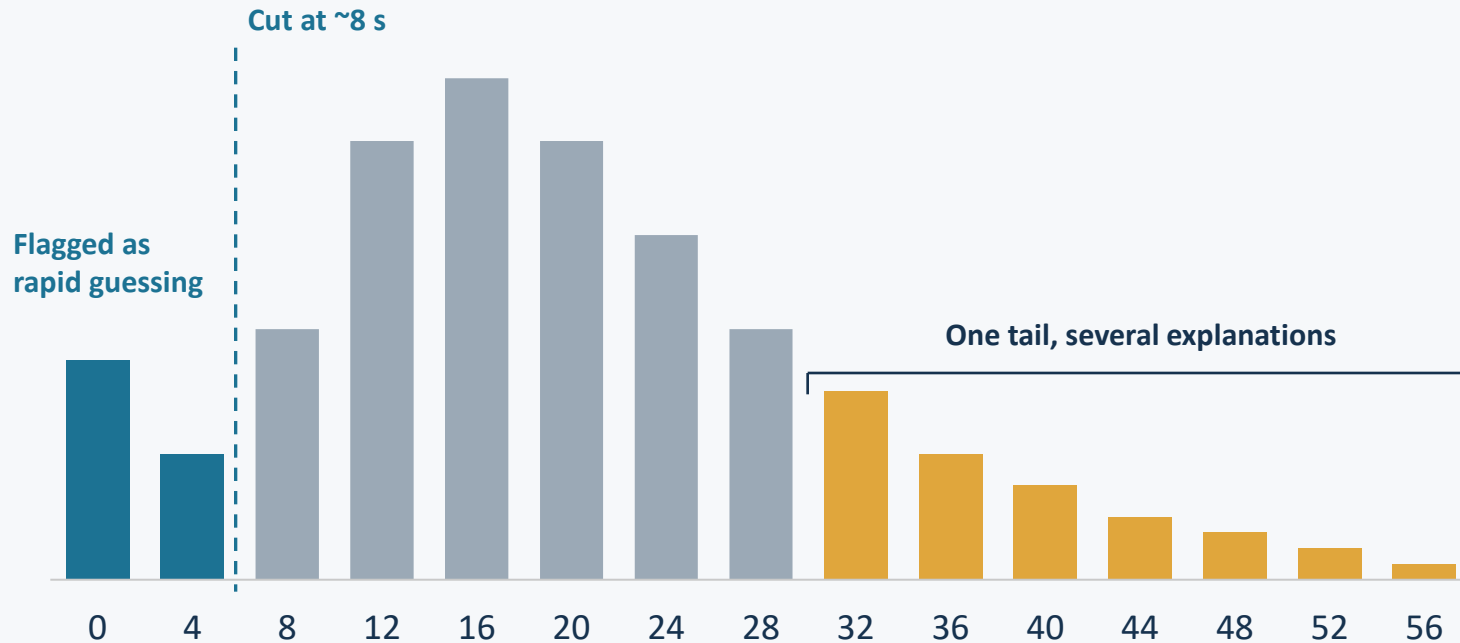
Pace & effort

Time-on-task as evidence



Thresholds work on the **fast** tail, not the **slow**

Response times on one item, distribution of responses, seconds



Fast tail

Under about 8 seconds, responses separate from the rest. A threshold flags rapid guessing with few false positives.

Slow tail

A 40-second response can be careful work, re-reading, struggling, or a learner who has walked away. The tail cannot say which.

A cut point is a measurement decision; the tail is an interpretation problem.

Note: Illustrative distribution, not observed data.



Strategy & structure

- Correct and incorrect problem-solving patterns recovered with Hidden Markov Models and network analysis ([Liu, Bulut, Cui, & Gao, 2024](#))
- Unsupervised anomaly detection in PIAAC problem-solving tasks: Aberrance without labels ([Bulut, Gorgun, & He, 2024](#))
- Behavioral patterns on interactive digital tasks using sequence mining techniques ([Ulitzsch, He, & Pohl, 2021](#))
- Web navigation patterns and problem-solving styles in interactive tasks ([Gao, Cui, Bulut, Zhai, & Chen, 2022](#); [Gao, Zhai, Bulut, Cui, & Sun, 2022](#))
- Writing process patterns using keystroke logs ([Zhang, Zhu, Deane, & Guo, 2017](#))

Each study presents a **highly plausible** explanation.

Methods in play

HMMs

latent process states

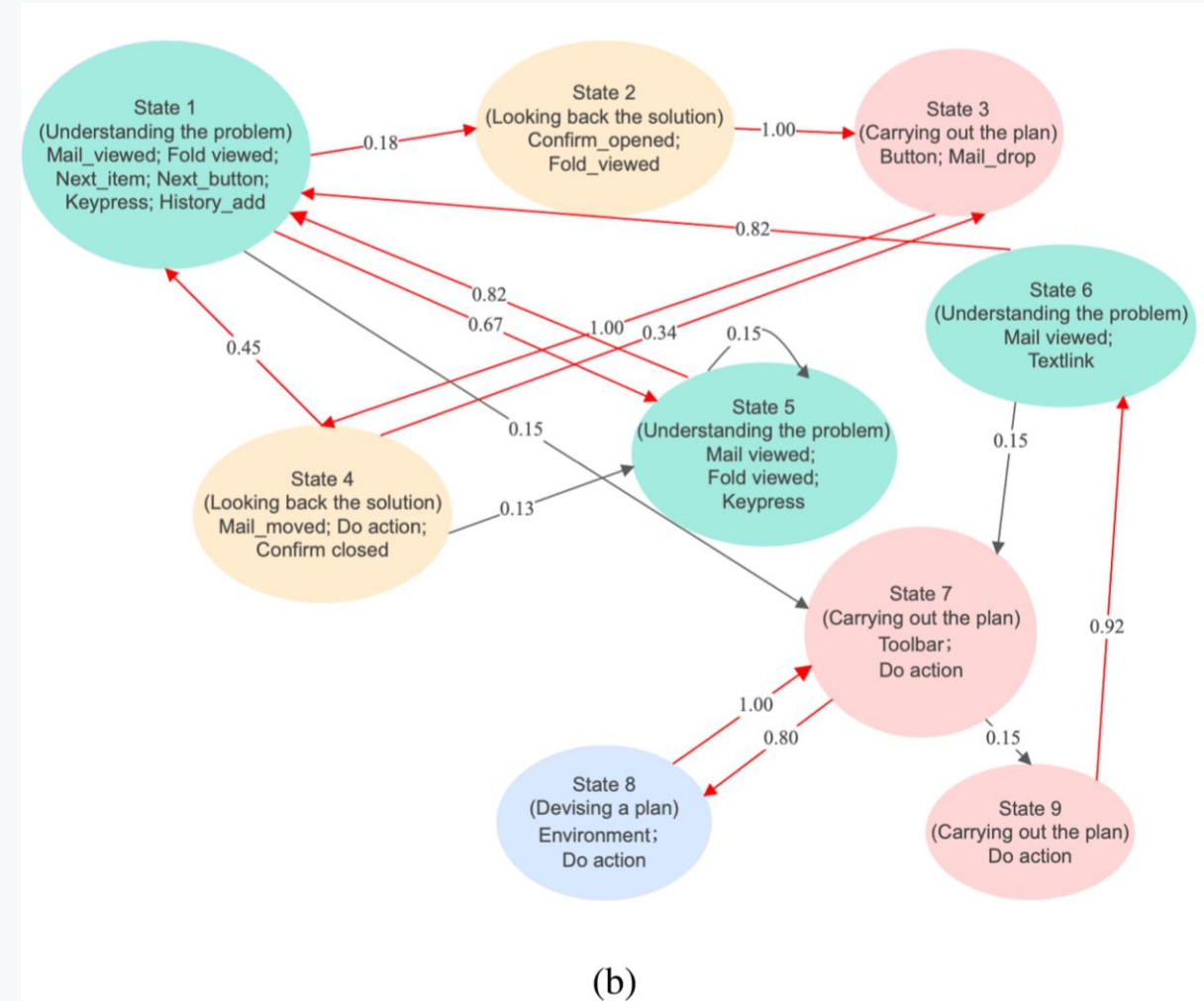
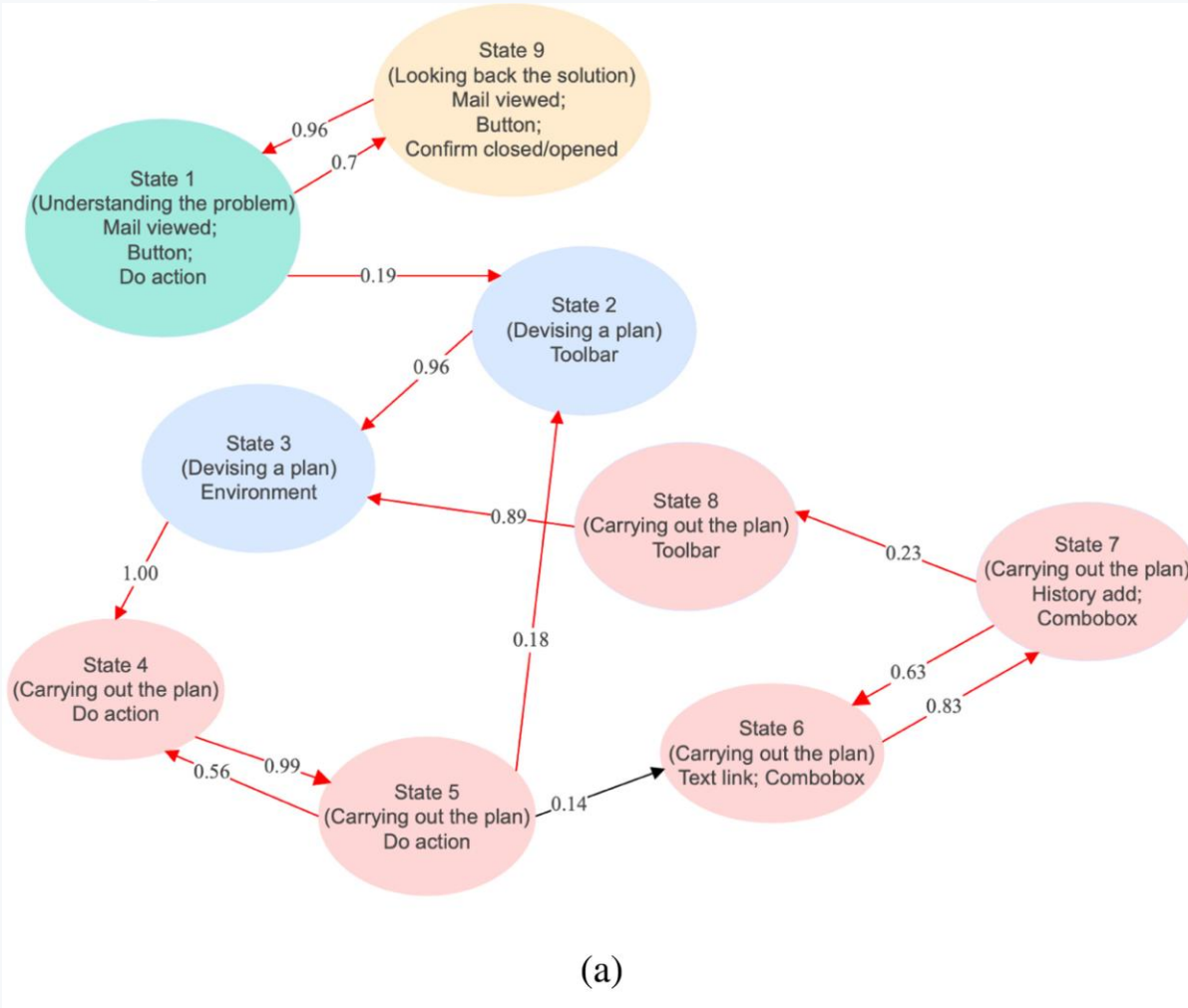
Networks

action co-occurrence

Anomaly detection

unlabeled, unsupervised

Problem-solving paths in the correct (a) and incorrect (b) groups for item U02 in PIAAC 2012 ([Liu, Bulut, Cui, & Gao, 2024](#))



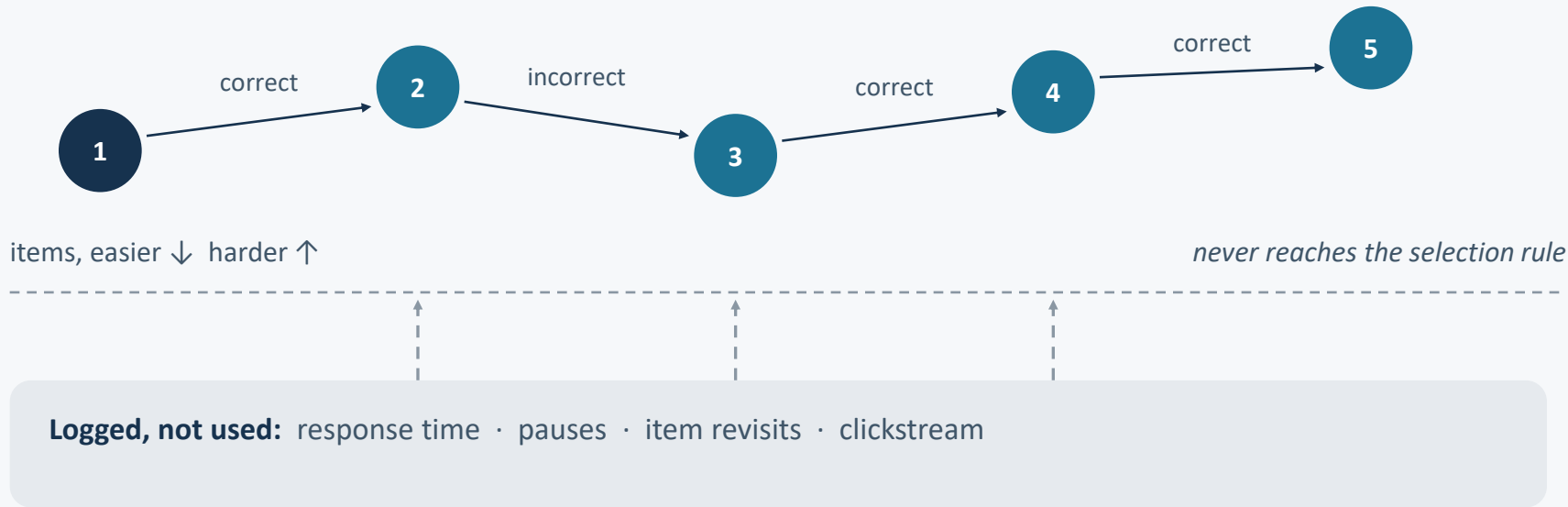
But explaining **mistakes** (rather than success) remains challenging.



Acting on process data (the weakest)

- Sequential modeling of responses and response times in a CAT with LSTM
- Adaptive item recommendation from process data and examinee background
- Creating value from process data: implications for multistage testing

How the next item is chosen: **correct or incorrect**



From signal to action

Predict

sequence → next state

Personalize

recommend the next item

Design

process-aware test forms

[Bulut et al. \(2023, NCME\)](#) · [Liu, Armoush, & Bulut \(2026\)](#) · [Bulut et al. \(2025\)](#)



Why interpretation remains hard



Response-time models admit multiple interpretations.

[*De Boeck & Jeon \(2019\)*](#)



Clickstreams are a partial, noisy record of activity.

[*Baker et al. \(2020\)*](#)



Log indicators need theory, context, and validation.

[*Goldhammer et al. \(2021\)*](#) · [*Zumbo et al. \(2023\)*](#)



The gap persists as machine learning advances.

[*Huang et al. \(2025\)*](#)



The foundation's unfinished business

We got good at **detecting** behavior.
Interpreting it stayed **hard**.

With the inspiration from Wilhelm Dilthey's [Die Natur erklären wir, das Seelenleben verstehen wir](#) (1894), we can claim that:

*Process data gave us **Erklären** (explain) and we are still reaching for **Verstehen** (understand).*

PART II

What's changing: from traces to talk

“Meaningful action can be read and interpreted as a text.”

—Paul Ricoeur, [*The Model of the Text*](#) (1971)





From low-semantic traces to high-semantic talk



TRACES · LOW-SEMANTIC

clicks · timing · keystrokes · tool use · action sequences

Tell you **THAT** it happened



TALK · HIGH-SEMANTIC

talk · chat · communication

reveals cognition, reasoning, collaboration

Tell you **WHY** it happened

Process data is acquiring intentionality!

[Maddox \(2017\)](#) · [Goldhammer & Zehner \(2017\)](#) · [Mohan, Bergner & Halpin \(2020\)](#) · [Zhang, He & Kwon \(2026\)](#) · [Chang & Li \(2026\)](#)



From action to intention

Conversational turns can be coded for the functions they serve, revealing engagement that behavioural traces alone hide.

● Clarification

resolving ambiguity

● Negotiation

aligning on approach

● Reassurance

managing confidence

● Strategy seeking

asking how to proceed

● Higher-order thinking

reasoning and evaluating

**+ other forms of cognitive
engagement**

[Bulut, Tan & Yildirim-Erbasli \(2026\)](#) · [Uzun, Gauthier & Cukurova \(2026\)](#) · [Chang & Li \(2026\)](#)



What this newly licenses

- Inferences about reasoning and intent, **not** merely observed behavior
- The sequential toolkit still applies, but now to semantic states rather than actions

Same toolkit, new questions it can answer

→ Process mining

Recover the dominant solution paths, and where reasoning departs from them

→ Lag-sequential analysis

Which state follows another more often than chance

→ Survival analysis

How long a misconception persists before it is repaired

→ Epistemic network analysis

Which concepts a learner connects, and how tightly

→ Trajectory clustering

Group learners by the route they reason through, not the score

→ Latent transition analysis

How state membership shifts from one occasion to the next

Transfers almost unchanged

→ Hidden Markov Models

→ Network analysis

→ Sequence models (LSTM)

The states just become semantic.



New power, old costs: In new form

1

Privacy

Utterances expose far more than clicks ever did.

2

Construct-irrelevant variance

Language fluency bleeds into the signal.

3

Fairness

Effects differ across linguistic backgrounds.

Semantic data do **not resolve these problems. It adds new ones.**

PART III

What breaks, and what to do





Two claims, different weight

THEORETICAL

The unit of analysis moves from the solver to the human-AI dyad.

Delegation, prompting, and negotiation aren't in our current models.

VALIDITY

Pre-AI findings may not transfer to a post-AI world.

The sharpest point for the measurement and assessment community: Comparability is now in question!



The solver becomes a dyad

- Existing theories assume the process is endogenous to the test-taker.
- Delegation, prompting, and negotiation restructure how the task is solved.
- We need theory grounded in **human–computer interaction (HCI)**.



solver



human + AI

the new unit of analysis

Double hermeneutic ([Giddens, 1976](#)): Social science interprets a world that is already interpreting itself.

Triple hermeneutic: We will interpret a learner who interprets an AI that interprets them.



Borrowing from HCI

The dyad already has a literature. Four ideas to build on:



Distributed cognition

The dyad is the unit of analysis.

[*Hollan, Hutchins & Kirsh \(2000\)*](#)



Common ground

Talk is how partners coordinate.

[*Clark & Brennan \(1991\)*](#)



Appropriate reliance

A measurable two-part ability.

[*Schemmer et al. \(2022\)*](#)



Human-AI teaming

Design for complementary performance.

[*Bansal et al. \(2021\)*](#)



A literature we should be reading

18 validated guidelines for human–AI interaction ([Amershi et al., 2019](#))

TO INSTRUMENT THE DIALOGUE

G11

Make clear why the system did what it did

Log the agent's reasons, **not** just its output.

G9

Support efficient correction

Repair and revision are high-signal process data.

ECHOES OF CONSTRUCT DRIFT

G14

Update and adapt cautiously

A system that shifts under the learner breaks comparability.

G18

Notify users about changes

Record when the task changed; version it like a norm.

Why this matters: Human–computer interaction has already studied how people work with AI. As AI enters the task, the assessment community also needs to pay more attention.

CONSTRUCT DRIFT

A long response time in 2018 and in 2026 are not the same event.

Comparability of pre- and post-AI process data becomes an open empirical question, not a working assumption.

Cognitive offloading: When learners shift work to the AI, the same trace no longer indexes the same internal process.

[Risko & Gilbert \(2016\)](#)

Two eras of process data.

Show

Hide

Run again

2016

2026

THE TASK, IDENTICAL IN BOTH ERAS

Which dial is which?

Two unlabelled dials control a sealed chamber. One sets temperature, one sets humidity. Change the dials, apply, read the gauges, then name each dial.

Dial 1

Low

Med

High

Dial 2

Low

Med

High

Apply

TEMPERATURE

22 C

HUMIDITY

50 %

ANSWER

Dial 1 = Temperature

Dial 2 = Humidity

Interaction trace

log_2016.csv · dial actions only

TIME	EVENT	DETAIL
00:00	Task loaded	Dial 1 Med · Dial 2 Med · Temp 22 · Hum 50
00:09	Dial 1 Med to High	
00:12	Dial 2 Med to High	
00:15	Apply	Temp 26 · Hum 70
00:19	[idle 41s]	no input
01:00	Dial 2 High to Med	
01:03	Apply	Temp 26 · Hum 50
01:11	Dial 1 High to Med	
01:14	Apply	Temp 22 · Hum 50
01:22	Answer submitted	Dial 1 = Temperature · Dial 2 = Humidity

Response times. Median action interval 3 s. One unexplained gap of 41 s at 00:19, with no record of what the student was doing.



No conversational agent available.

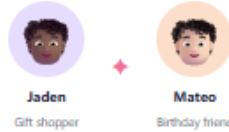
The 2016 platform captured dial actions only. Whatever reasoning happened during the pause left no record.

Find a gift that says, “I get you.”

Help Jaden choose a birthday gift for Mateo. The gift needs to fit the budget, match Mateo's interests, and arrive before the party.

PARTY COUNTDOWN
🕒 **4 days, 6 hours**
Saturday - 2:00 PM

01 Meet the friends



Jaden wants to get Mateo something thoughtful—not just something that looks cool.

Budget: \$45
Including delivery

Likes space & games
Also enjoys drawing

Party is Saturday
Needs to arrive by Friday

Coach tip Ask Jaden what Mateo likes before you recommend a gift.

CONVERSATION

Talk with Jaden

4
turns

week he was showing me a planet sketch he made. I want the gift to feel personal.

00:16:31 AM

When is the party, and when does the gift need to arrive?

YOU

00:16:33 AM

The party is Saturday at 2 PM. I need the gift by Friday night so I can wrap it. The headset arrives Saturday, which is too late. The other options arrive from Wednesday to Friday.

00:16:33 AM

Which two gifts should we compare first, and what should we look at?

YOU

00:16:34 AM

Let's compare two gifts side by side. Click a gift on the shelf to open it, then use "Add to compare." Add one more gift, then press Compare. Check three things: total cost ($\leq$ \$45), arrival (by Friday), and interest fit.

00:16:35 AM

Ask about Mateo

Check the deadline

Compare two gifts

Ask Jaden a thoughtful question...



02 Gift shelf

Open a gift for details. Then add two gifts to the compare tray.

- Star Explorer Kit**
\$34 + free delivery · Fri. →
- Nova Gaming Headset**
\$42 + \$6 delivery · Sat. →
- Galaxy Sketch Set**
\$28 + \$5 delivery · Thu. →
- Pixel Planet Lamp**
\$39 + free delivery · Thu. →
- Cosmic Card Game**
\$24 + \$4 delivery · Wed. →
- Orbit Hoodie**
\$37 + \$4 delivery · Fri. →

Star Explorer Kit

TOTAL COST
\$34

DELIVERY
Friday

INTEREST FIT
Strong match: Mateo loves space

WHAT'S INSIDE
Small telescope, star map, and planet cards.

Add to compare

COMPARE TRAY
Choose two gifts to compare

Compare →

Make a recommendation →

YOUR THINKING TRACE

How you're solving the mission

3

smart questions

4/4

facts explored

97%

collaboration

View activity log ↗



A tentative agenda: Four footholds

From concrete instrumentation to a field-level ask.



ascending scope →



From dialogue traces to a theory of the dyad

1 Instrument for dialogue, not just traces

- Log prompts, responses, revisions, and inter-turn timing
- Annotate turns by intent: asking, verifying, delegating, rejecting

2 Model the dyad

- Borrow from HCI: distributed cognition, common ground, reliance/calibration
- **Candidate construct:** Effective use of an AI collaborator



Anchor the comparisons, and share the instruments

3 Make comparability testable

- Invariance-style tests across pre- and post-AI cohorts on process features
- AI-free anchor tasks that hold a fixed reference point

4 Build shared reference tasks

- Common instrumentation standards across platforms and studies
- So, results accumulate instead of fragmenting per platform



An invitation, not a verdict

Not a ruling on what to retire: an agenda the field can take up together.

● Continuity

The methodological toolkit endures

● Change

The data learns to talk

● Comparability

Held open, and testable

The world has already changed, and that is
the **invitation** for all of us.



Thank you

Questions & discussion

Okan Bulut

Measurement, Evaluation, and Data Science · Faculty of Education
University of Alberta

bulut@ualberta.ca · www.okanbulut.com

