Large-Scale Research on Engineering Design Based on Big Learner Data Logged by a CAD Tool

Charles Xie, Saeid Nourian, & Helen Zhang, Concord Consortium; Şenay Purzer & Robin Adams, Purdue University

How big will our data be?
- 3,000 students
- 20,000 hours
- 60 gigabytes

How will we collect these data?
What will we do with these data?
What will we find from these data?

Three types of structured data streams that record all student actions, artifacts, & articulations (AAA)
How will we collect these data?

Energy3D: The ONLY computer-aided design (CAD) software that logs “atomically” fine-grained process data about what students do, make, and say — behind the scenes.

Testbeds: Solar Urban Planning, etc.

“A Big Data Microscope”

STEM areas: Geometry, energy, thermodynamics, heat transfer, architectural engineering, building science, renewable energy, sustainability, ... (NGSS MS-PS3-3/4 and HS-PS3-1/3/4)

To probe how students learn in great details, we must build educational versions of “large hadron colliders!”
What will we do with these data?

Data are “mind recorders”: Use data as METHODS, not just OUTCOMES!

- Define questions
- Design research
- Collect data
- Analyze data

Learning dynamics visualization: Complete reconstruction and visualization of engineering design processes

- Time series analysis
- Signal processing
- Pattern recognition
- Machine learning

(4,000-6,000 actions 300-500 artifacts, & 500-1,000 words per student)

Computer logs are instructionally sensitive!
What will we find from these data?

• What are learning trajectories and patterns that characterize iterative design, problem space exploration, convergent-divergent thinking, creativity, novice-expert transition, gender differences, etc.?

• What are the bottlenecks that block knowledge and skill transfer? What makes it difficult for students to learn and apply science concepts in engineering design projects (Vattam & Kolodner’s “design-science gap“)?

• What are the “chemical reactions” and “phase change” in learning dynamics? How do we find evidence of them from data? How do we engineer those cognitive processes pedagogically (and test it)?

A HUGE array of questions to ask and explore!