The Concord Consortium

AERA 2016, Washington DC

Visualizing Engineering Design Processes of High School Students Using a CAD System

Presenter: Charles Xie

Contributors: Corey Schimpf, Jie Chao, Saeid Nourian, Guanhua Chen

The Concord Consortium

This work is supported by the National Science Foundation (NSF) under grant numbers 1348530 & 1503196. Any opinions, findings, and conclusions or recommendations expressed in this material, however, are those of the authors and do not necessarily reflect the views of the NSF.

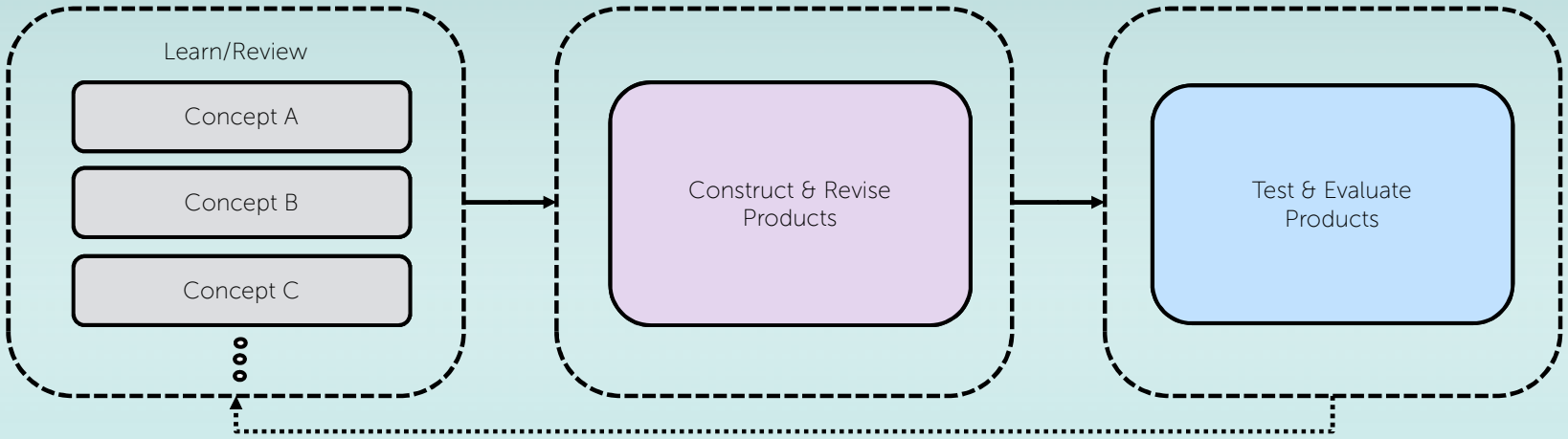
The big picture

So why are we doing this?

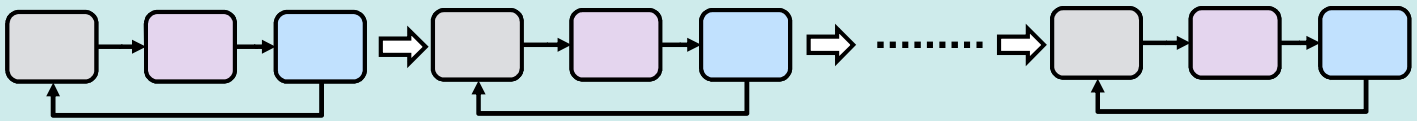
Engineering design *is* a complex process.

Summative assessment based on analyzing final products is too little too late. We want to help students while they are on it. Continuous formative feedback is a key.

Common Approach
Coarse-Grained, Delayed Feedback

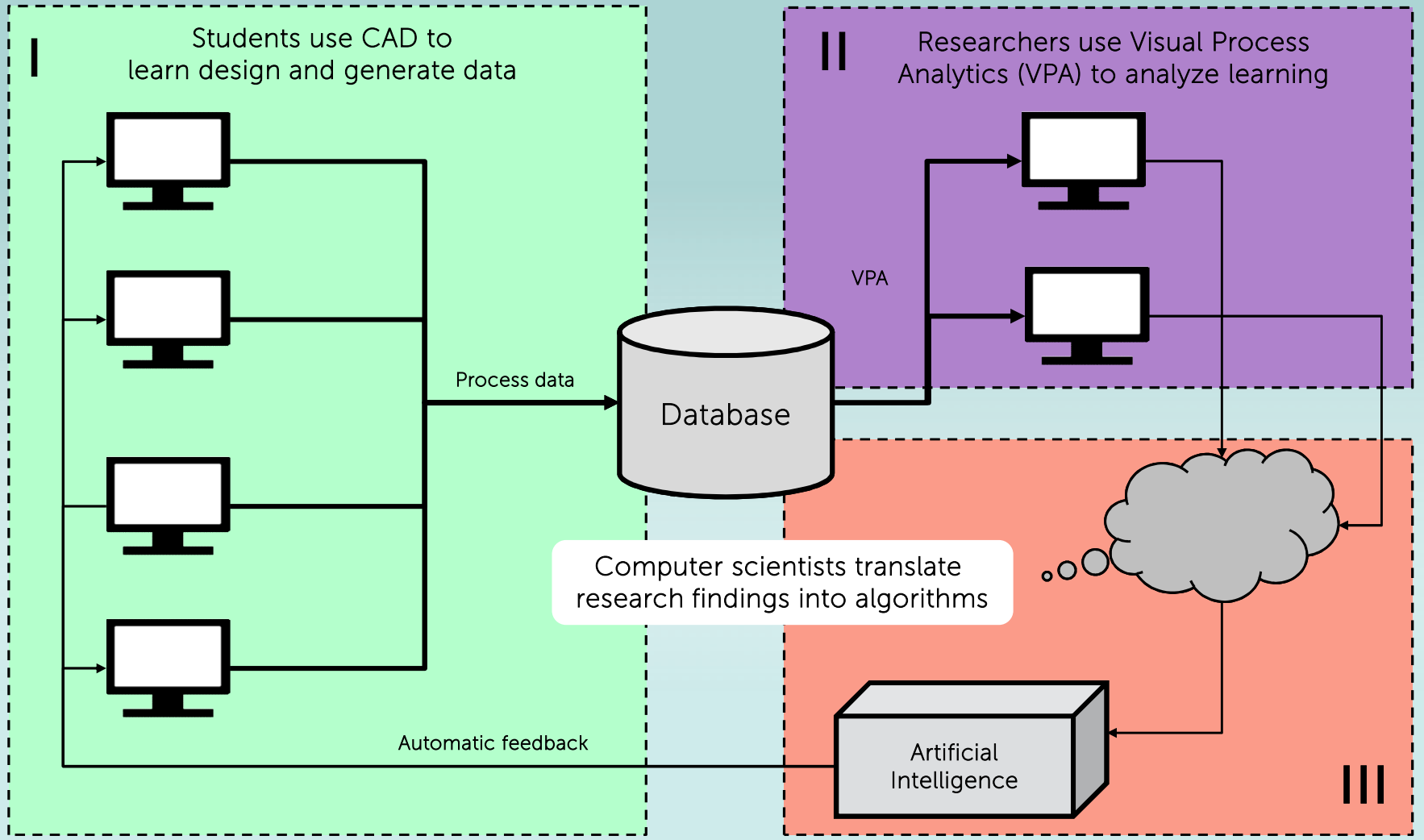


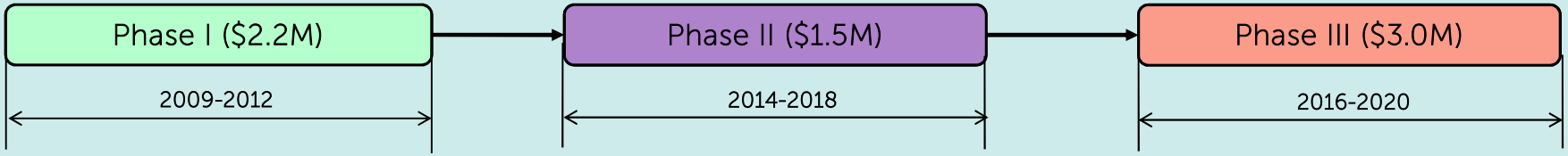
New Approach
Fine-Grained, Real-Time Feedback



Teachers can't look after every student at every second. Fair analysis of complex design may also be too time-consuming as it needs to include numerous factors. But computers may be able to do that!

The SmartCAD Vision



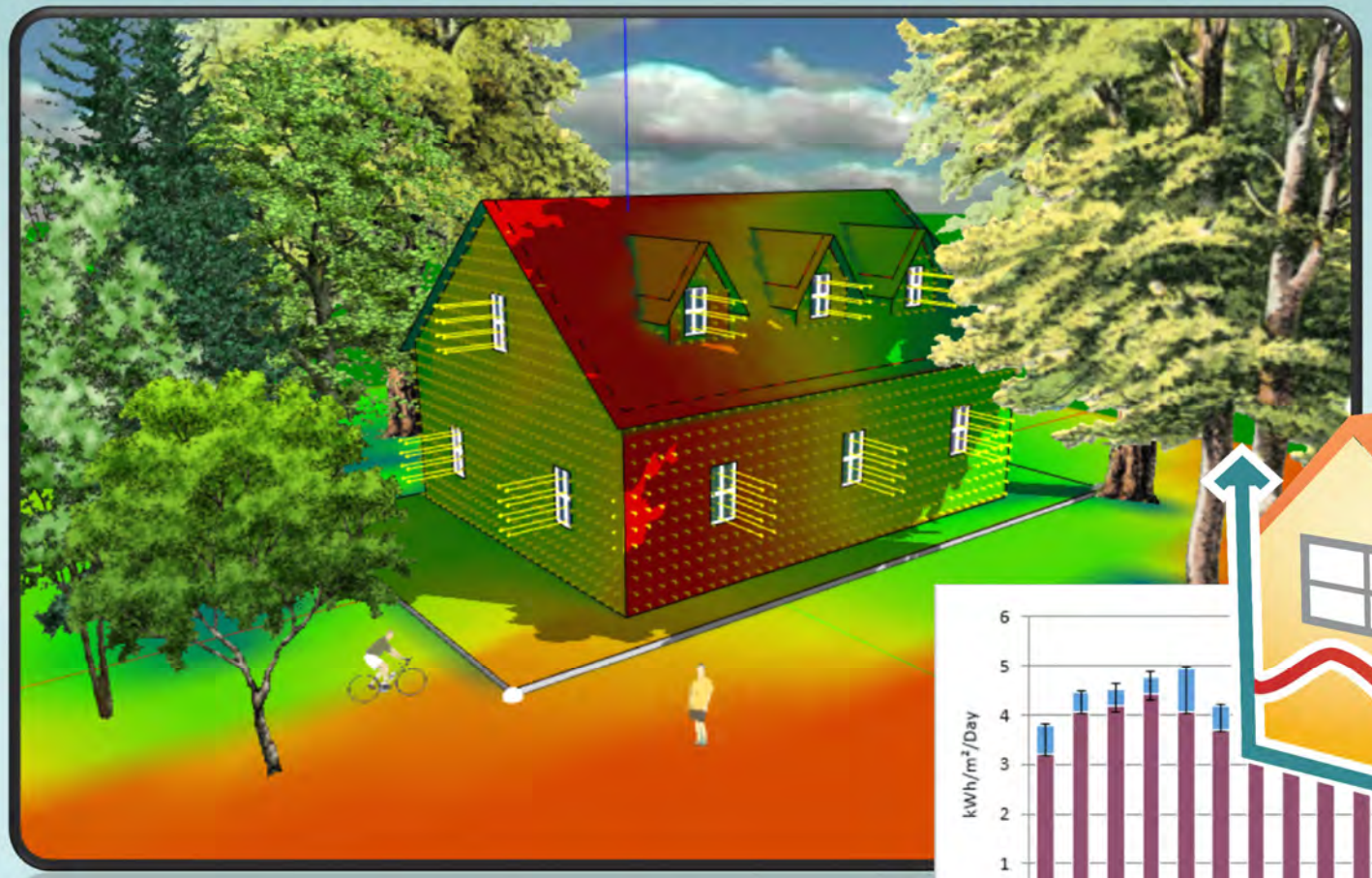


NSF: DRL-0918449, DUE-1348530, & DRL-1503196, PI: Xie

The research platform

The Energy3D SmartCAD program: A simulated engineering design environment

(An open lab for anyone to conduct data-intensive research on engineering design)



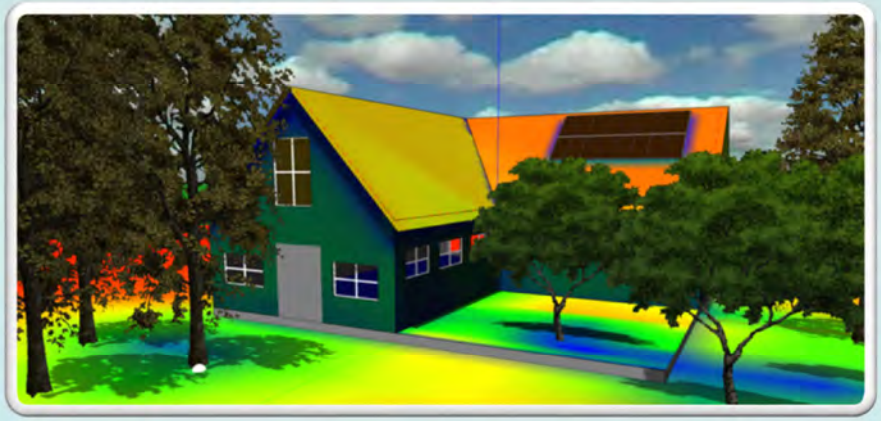
- Architectural engineering
- Solar engineering
- Energy engineering
- Urban planning
- ...

We strive for serious engineering precision and predictive power!

http://energy.concord.org/energy3d

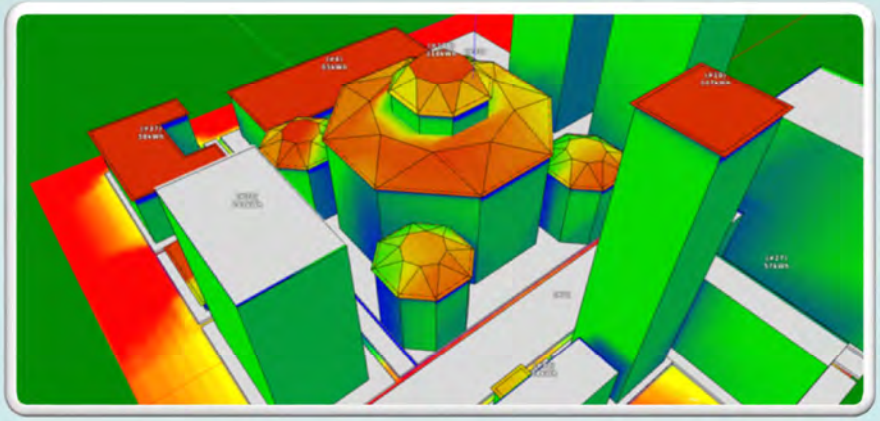
Research Subjects and Settings

Year	#students	Class	Grade	State	Design Challenges
2012	20	Engineering	Mixed	MA	Solar Urban Design
2013	63	Engineering	Mixed	MA	Solar Urban Design
2013	68	Physics	9	MA	Solar Urban Design
2014	67	Physics	9	MA	Energy-Plus Home Design
2015	110	Physics	9	MA	Energy-Plus Home Design + Solar Urban Design
2016	70	Physics	Mixed	MA	Energy-Plus Home Design
2016	70	Geoscience	Mixed	MA	Energy-Plus Home Design



Energy-Plus Home Design

Design a house that generates more renewable energy than it consumes over the course of a year

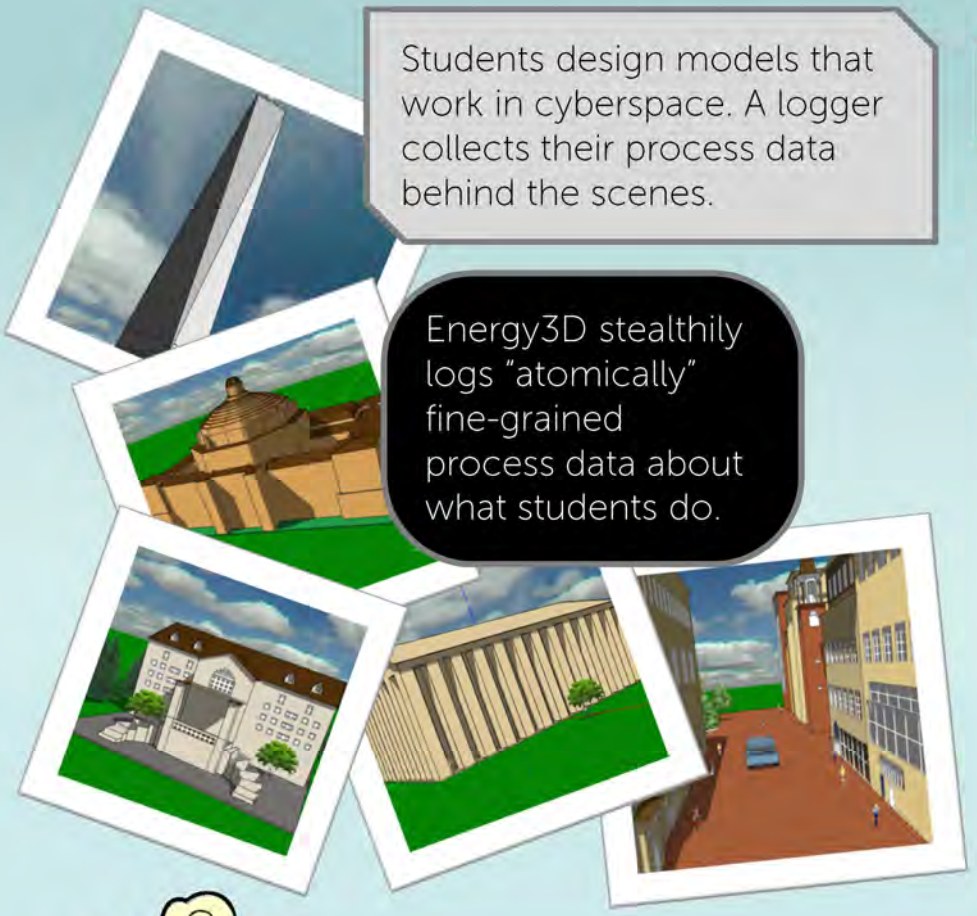


Solar Urban Design

Design a city block with high-rise buildings that have optimal solar gains in different seasons

The data

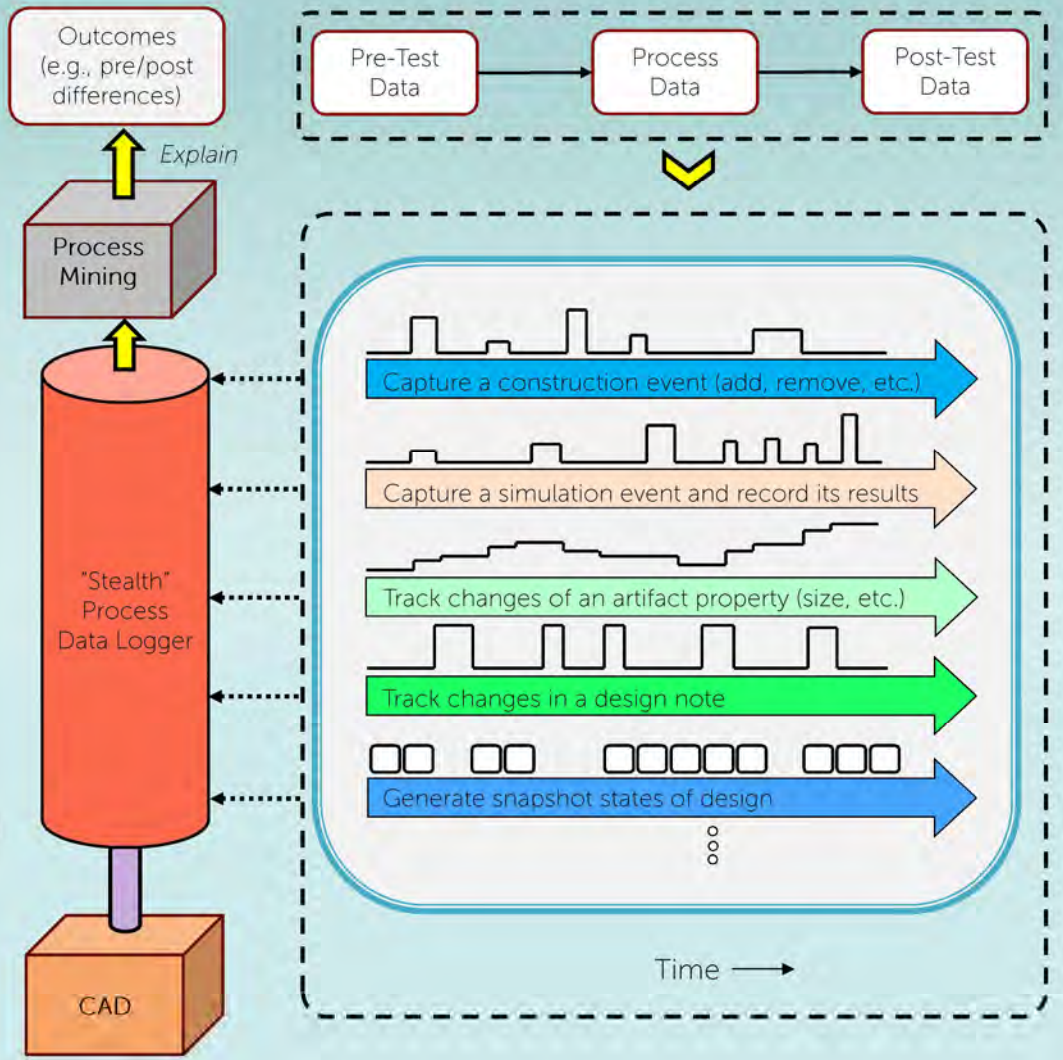
Collecting empirical "atomic" process data



Students design models that work in cyberspace. A logger collects their process data behind the scenes.

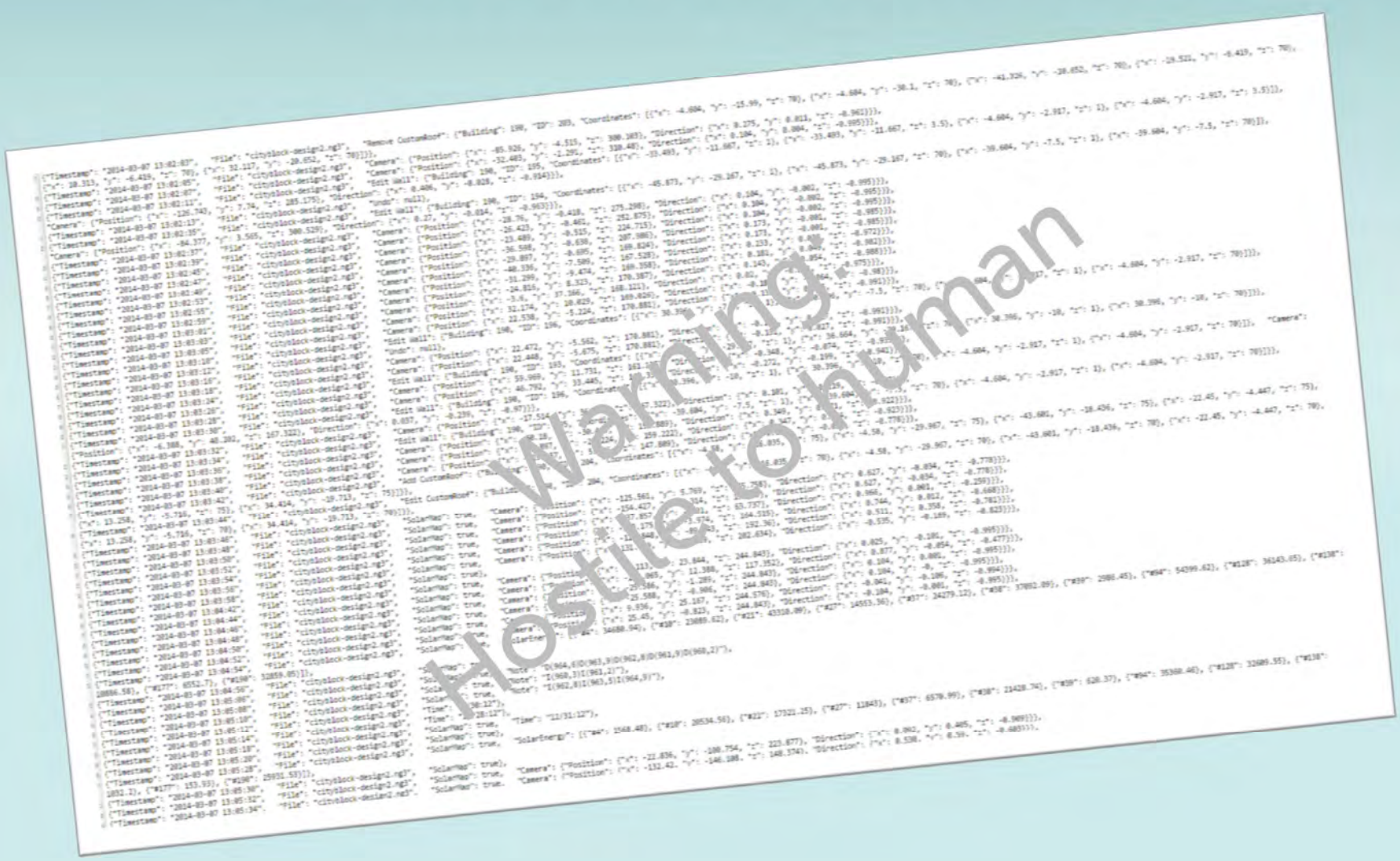
Energy3D stealthily logs "atomically" fine-grained process data about what students do.

Our invention: A data logger based on Undo Manager and Track Change (more structured than mouse clicks)



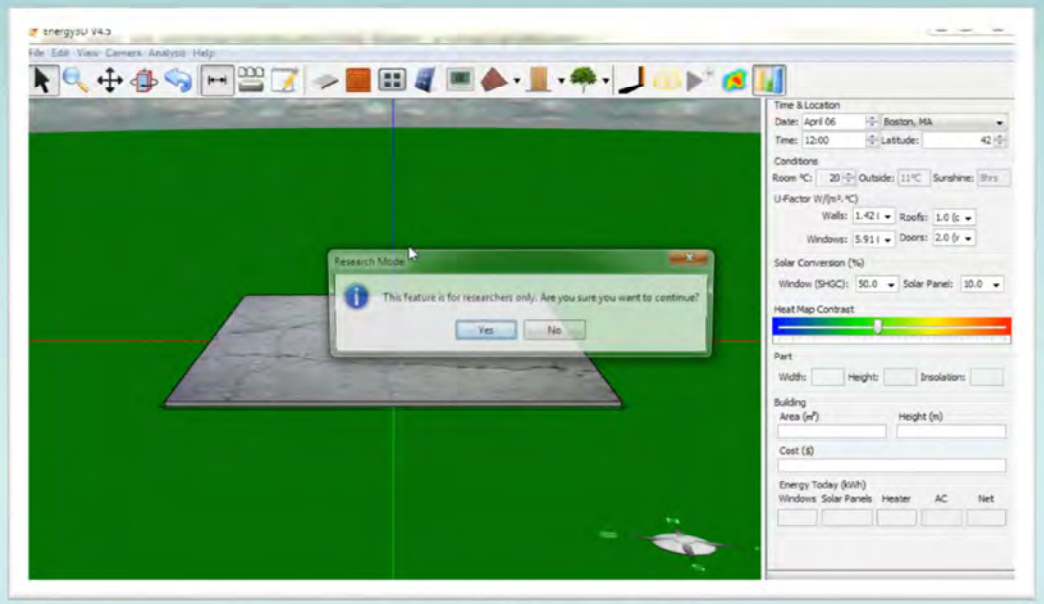
A JSON data schema encoding Energy3D design processes

How does the raw data stream look like?



Design replay

Reconstruct a design process from the data log and
play it back like running a slide show and post-process it to extract information as needed



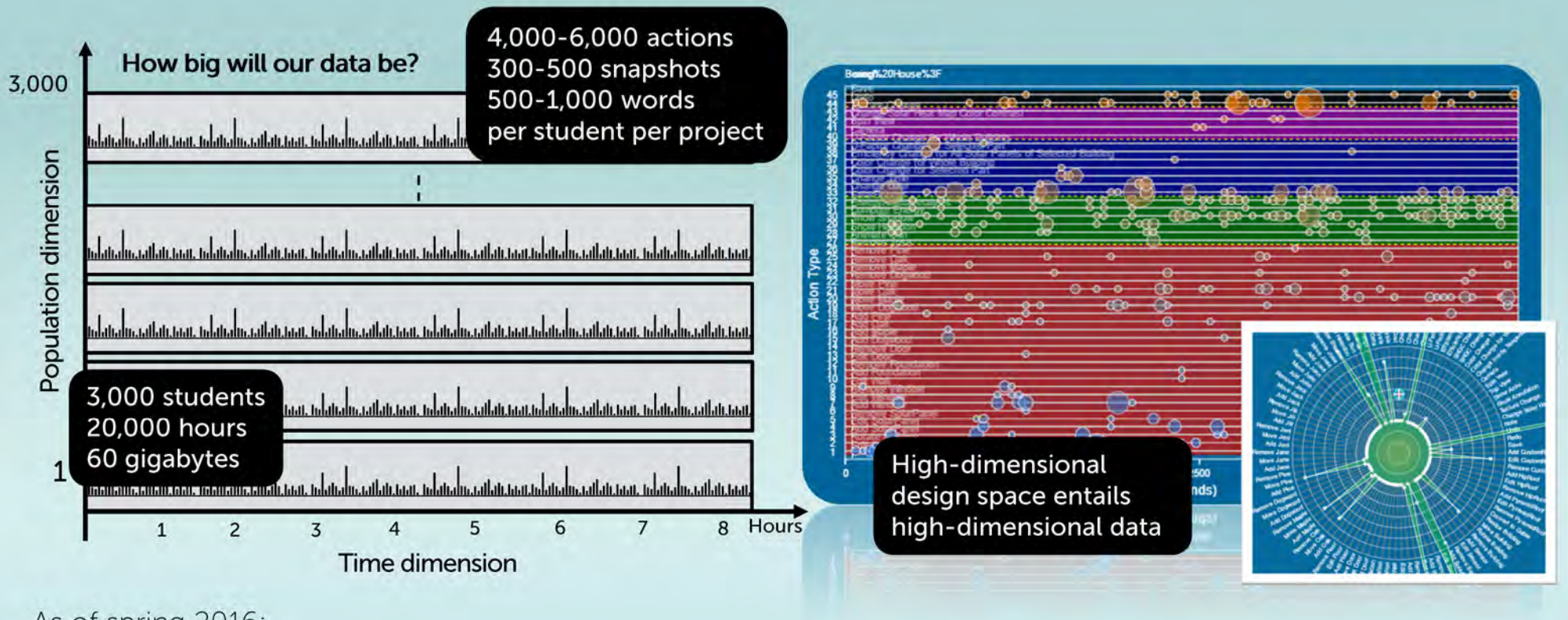
(Embedded video)

Compare with screencast, recording is based on events, not lapse of time.
(i.e., no event, no record.)

High ratio of lossless compression!

Data-intensive research

(aka "big data" – the fourth paradigm of science)



As of spring 2016:
1,000+ students' data will be in our repository.

What can we do with these data?

What can we find from these data?

Data mining

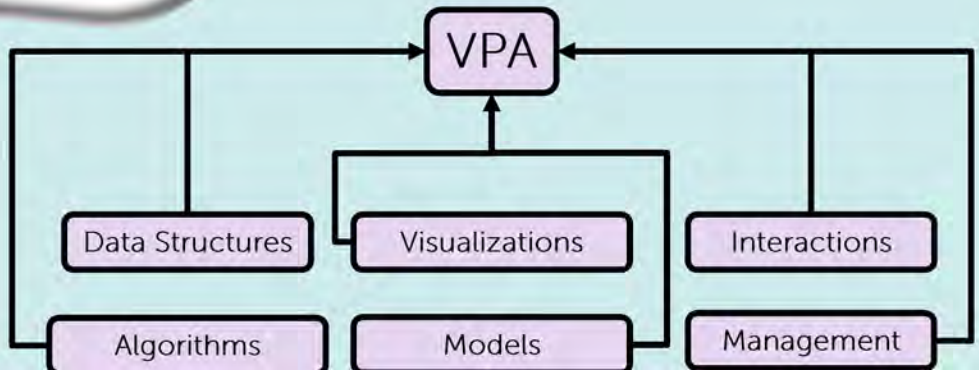
Visual Process Analytics (VPA)

Let's start with visualizing the data.

VPA is a Web-based data mining platform that supports research on student learning through using complex tools to solve complex problems.

http://vpa.concord.org

- Process mining
- Process modeling
- Process discovery
- Conformance analysis
- Statistical analysis
- .....



VPA supports multiple representations of data

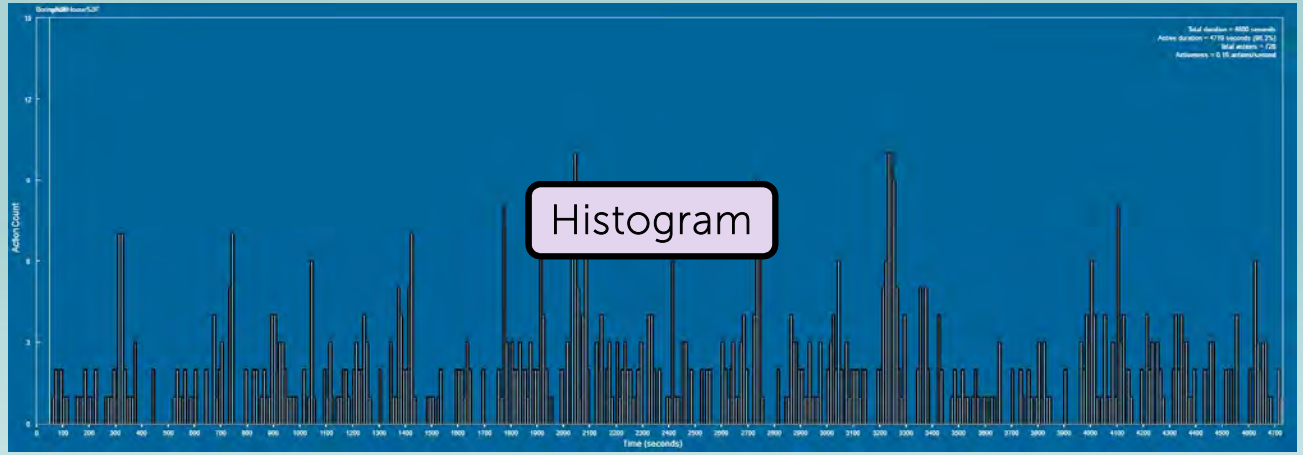
Time series visualization

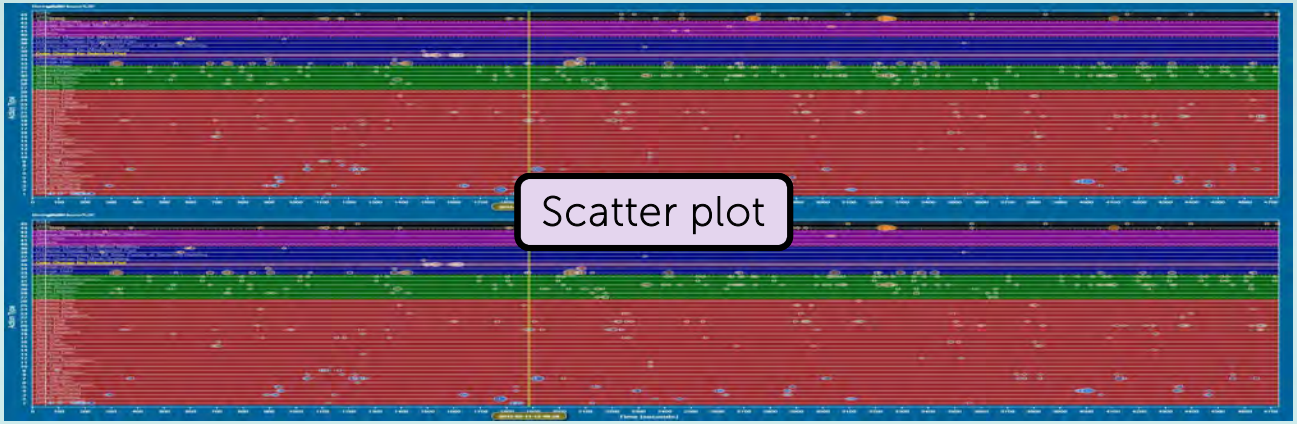
Histogram shows the total number of actions within each time bin

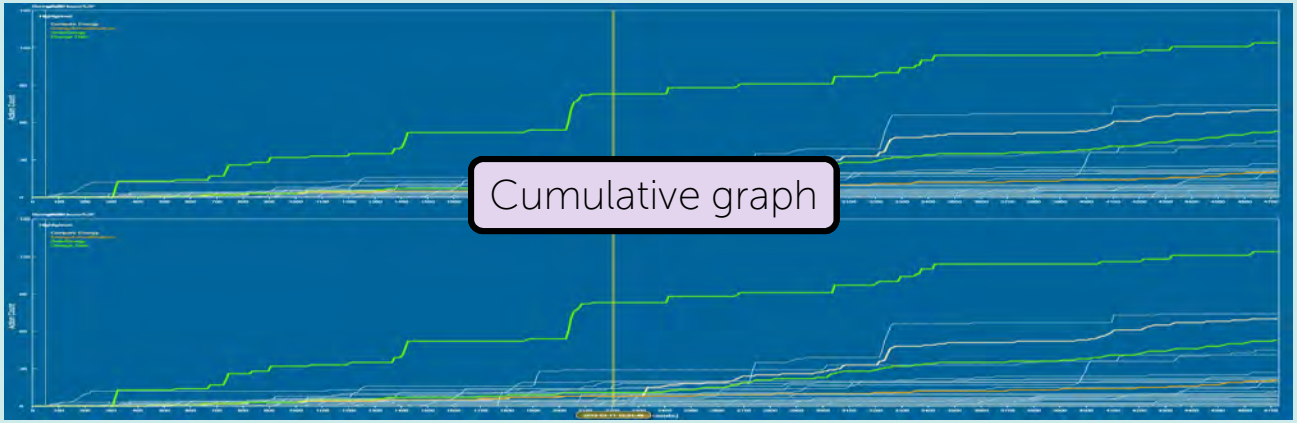
Scatter plot shows the number of actions of different types within each time bin

Cumulative graph shows the growth of the total number of actions of different types

Each kind of visualization represents a different view of the data and a different aspect of the process.

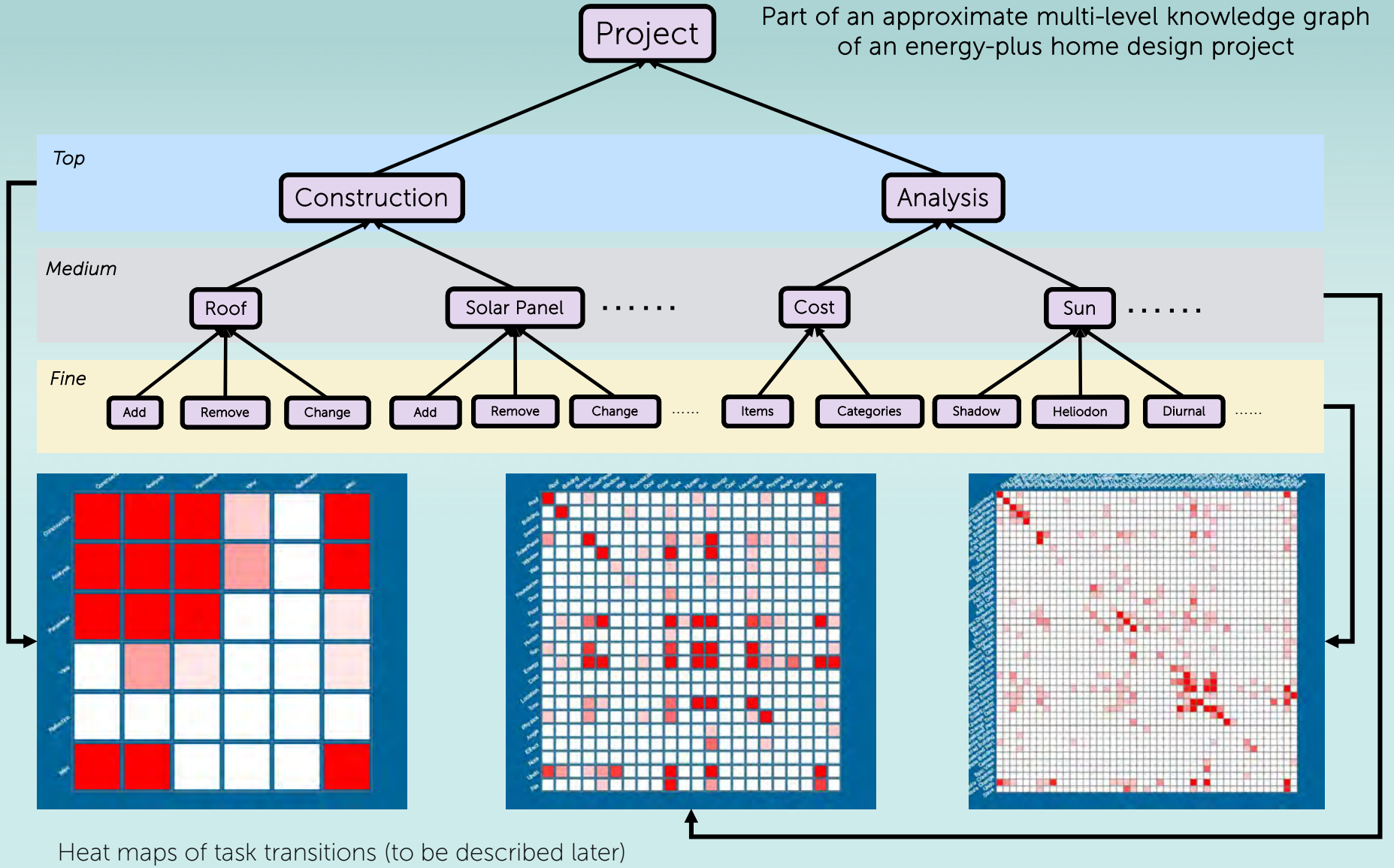






VPA supports multiple granularity of visualization

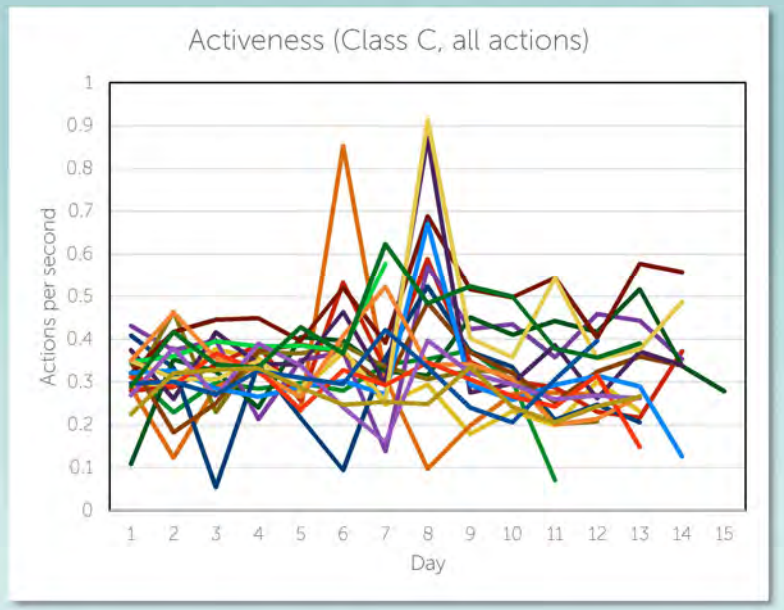
Coarse-grained vs. fine-grained analysis across knowledge graphs

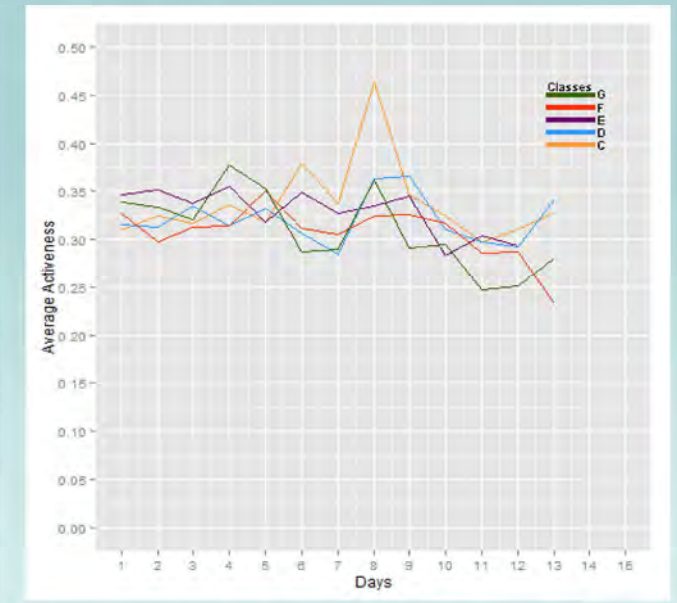


Heat maps of task transitions (to be described later)

VPA supports data collection and export

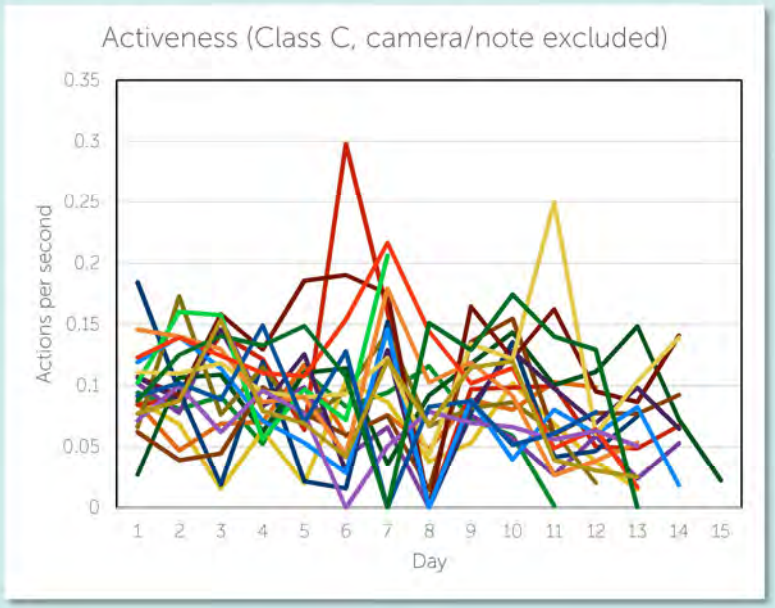
Mined results can be collected and exported for further analyses.

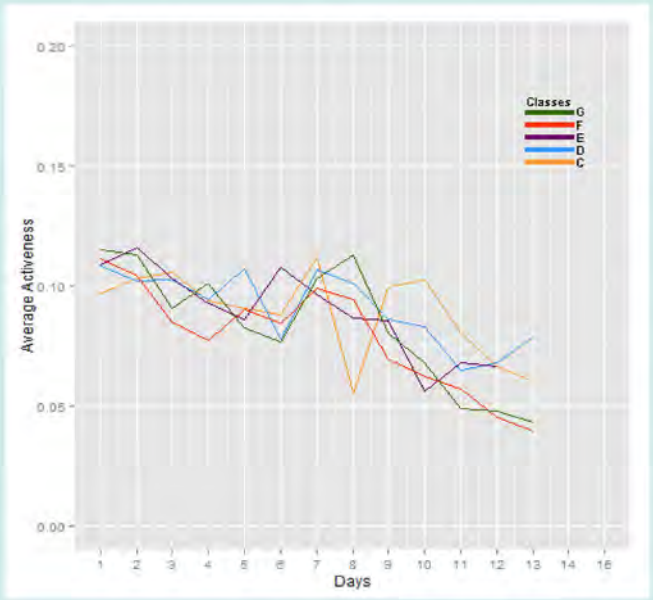




The change of the average activeness of five classes of students over time

(a total of 110 students)





The change of average activeness of five classes of students after filtering out 3D rotation and note taking, showing a gradual decrease over the course of two weeks.

Excel

RStudio

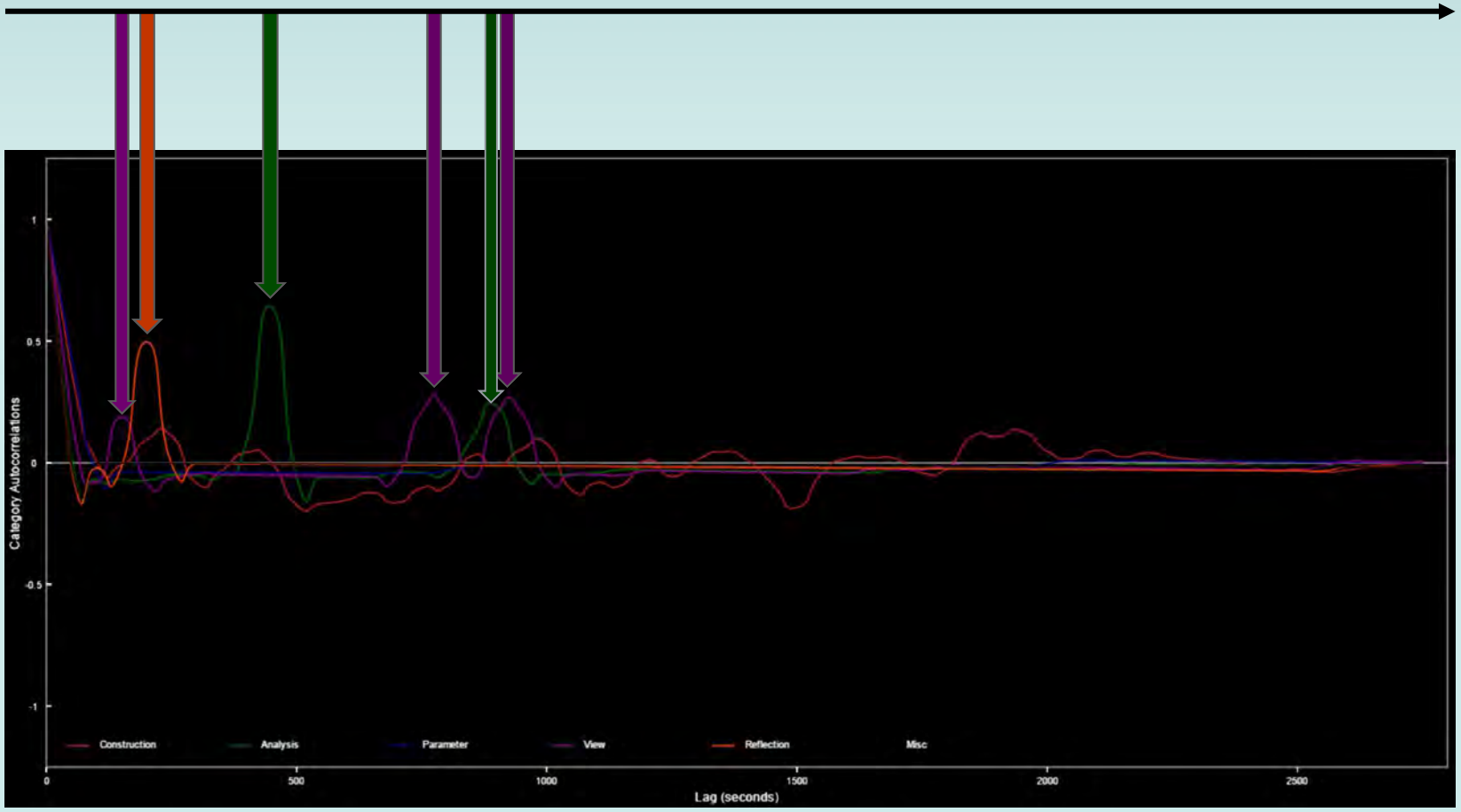
VPA tools

VPA tool: Time correlation functions

Correlograms show repeating patterns of behavior: After how long, on average, do students repeat certain types of actions (an indicator of possible design iteration)?

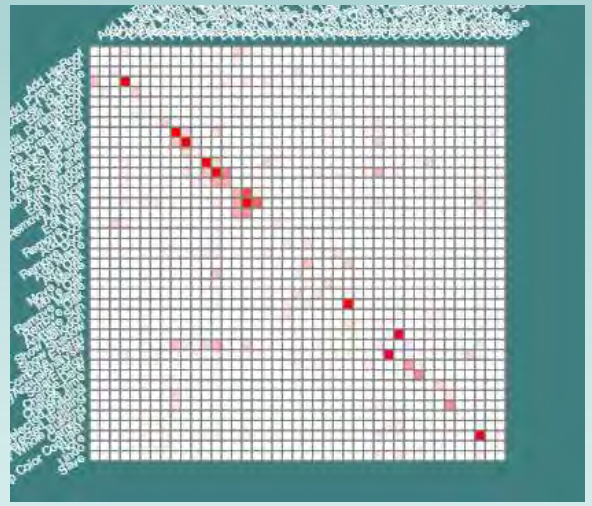
Construction (~200s), analysis (~450s)

$$(f \star g)(\tau) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f^*(t)\ g(t+\tau)\,dt,$$

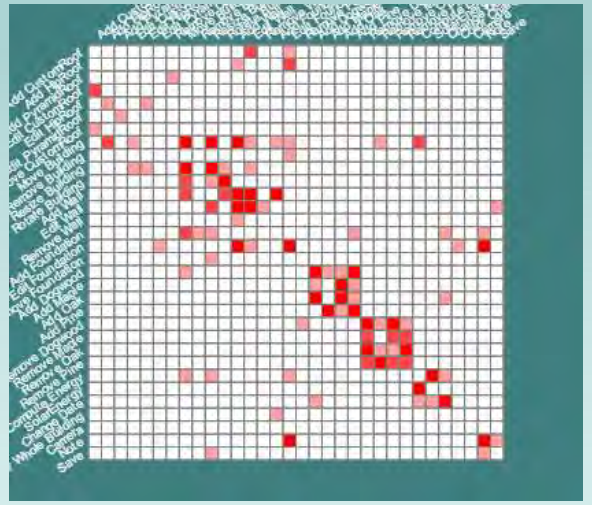


VPA tool: Heat map of task transitions

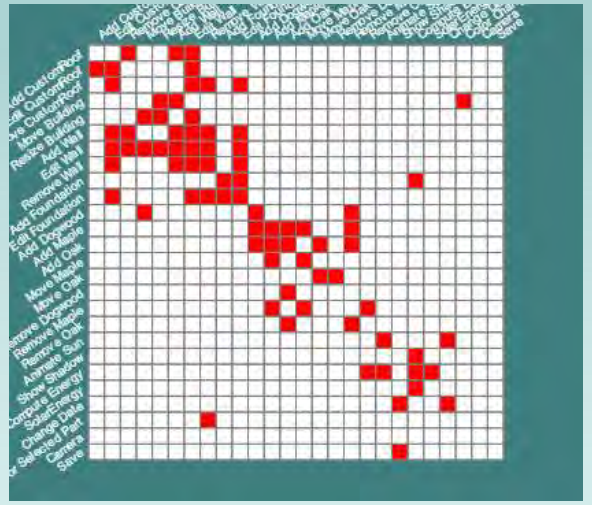
Transitions from tasks to tasks may reflect how students use the CAD tool to solve a design challenge.



Few transitions



Localized transitions



Frequent transitions

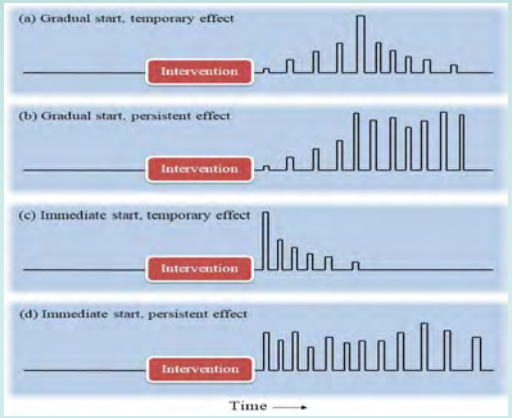
* This kind of heat map is a visual representation of the adjacency matrix of a design graph.

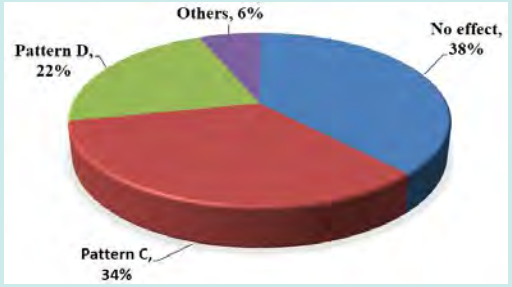
VPA tool: Response functions

How do students respond to an intervention?

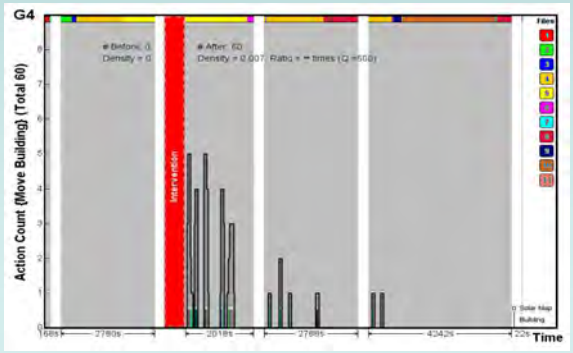
(An intervention can be computer-generated feedback, teacher instruction, or student discussion.)

$$x(t) = \int_{-\infty}^{t} R_x(t-\tau) I(\tau) d\tau$$

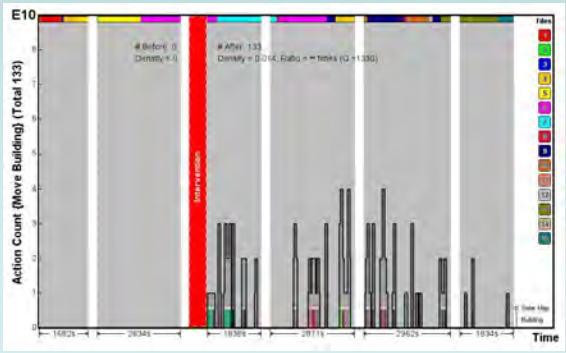




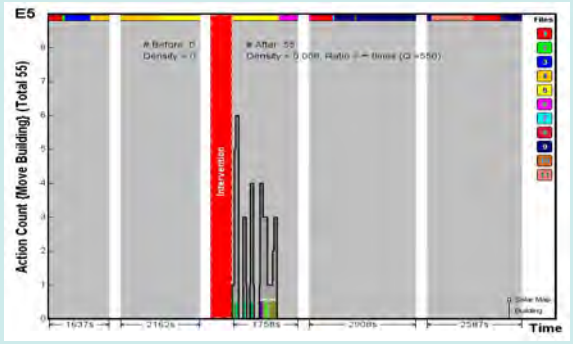
The distribution of response patterns of 65 students

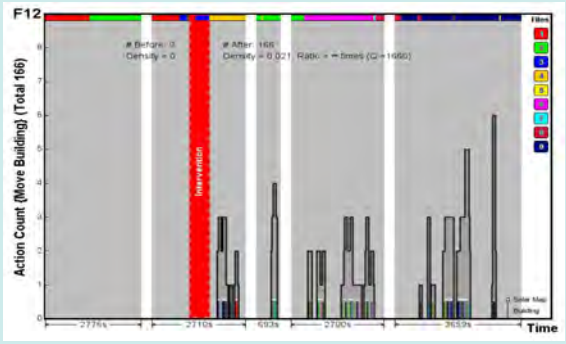


Pattern C: Decay



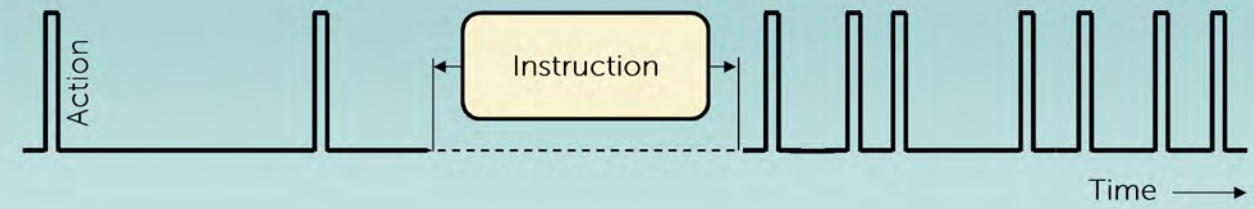
Pattern D: Persistent



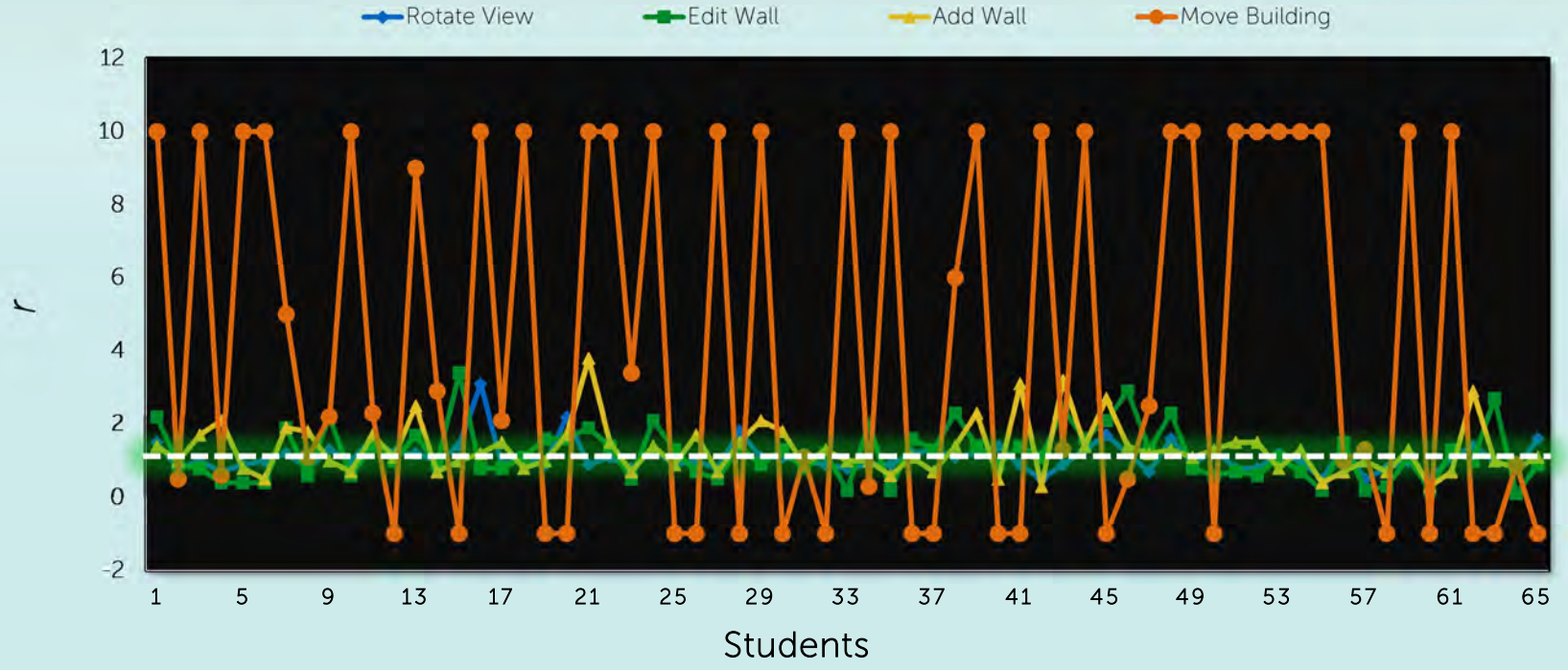


VPA tool: Response functions (cont.)

Comparing the changes of actions related and unrelated to a specific intervention



Ratios of Post/Pre-Instruction Action Densities $r(A) = \frac{N_{\text{after}}(A)/T_{\text{after}}}{N_{\text{before}}(A)/T_{\text{before}}}$



Conclusion

Fine-grained process data in the CAD log encode the dynamics of engineering design supported by the CAD tool and regulated by external factors. As design is an open-ended task in a high-dimensional problem space, these data appear to be highly irregular, making them extremely difficult to analyze.

VPA provides a "data microscope" for researchers to get a sense of the "shapes of data" rapidly. Combining the computational power of the machine and the pattern recognition power of the brain, VPA is a tool for tackling the "big data" challenge.

A series of research papers based on VPA analysis of engineering design will be submitted for publication later this year. VPA data repositories and analytic tools are freely available online to anyone who is interested in studying engineering design.

Thank you
for your time!