

2018

Continuous improvement: How systems design can benefit the data-driven design community

Lomas, James and Patel, Nirmal and Forlizzi, Jodi

Suggested citation:

Lomas, James and Patel, Nirmal and Forlizzi, Jodi (2018) Continuous improvement: How systems design can benefit the data-driven design community. In: Proceedings of RSD7, Relating Systems Thinking and Design 7, 23-26 Oct 2018, Turin, Italy. Available at <http://openresearch.ocadu.ca/id/eprint/2704/>

Designing for Continuous Improvement in K12

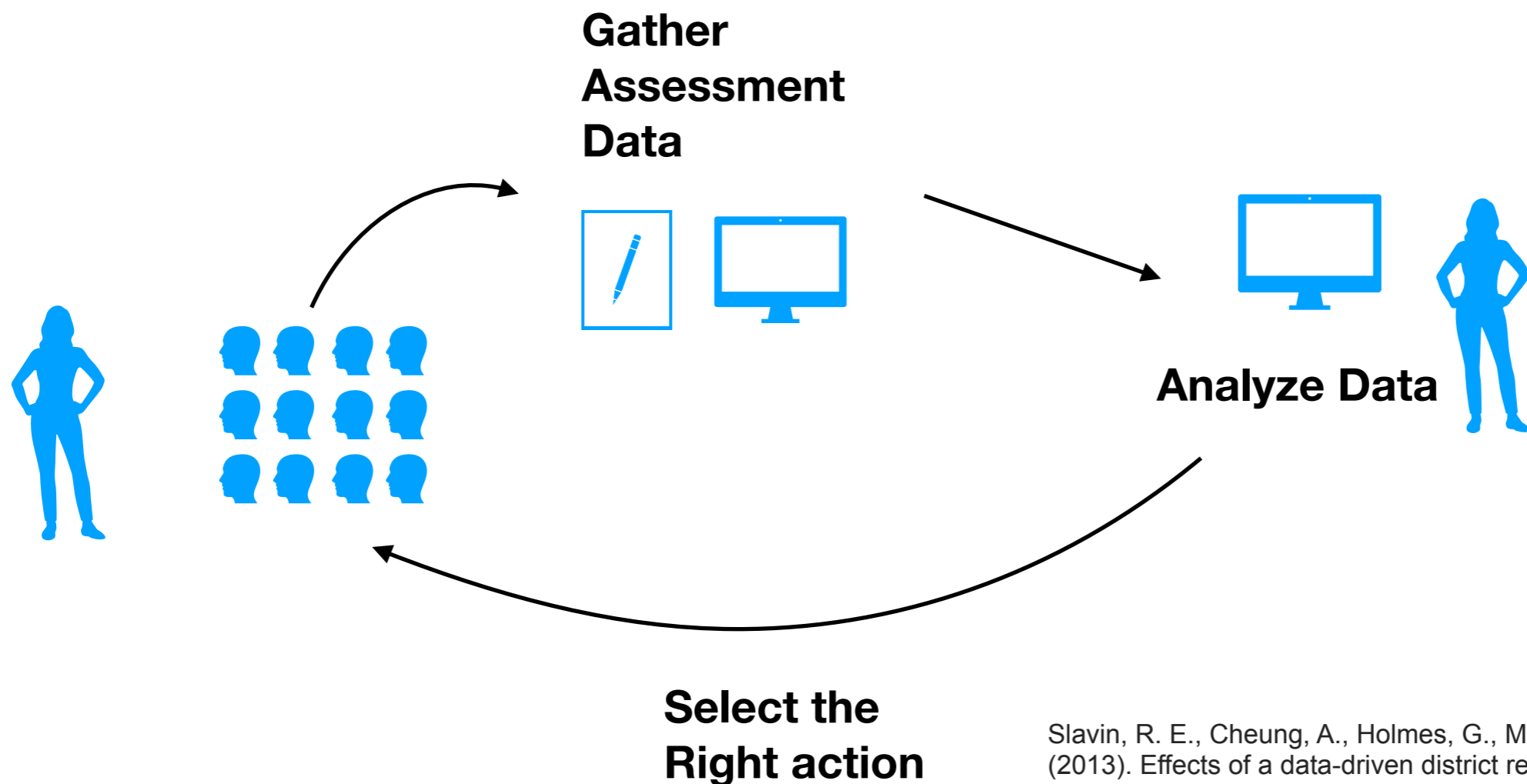
How systems design can help data-driven design in
complex-technical systems

Data-Driven Continuous Improvement in K12

"Data-Driven Decision-Making"

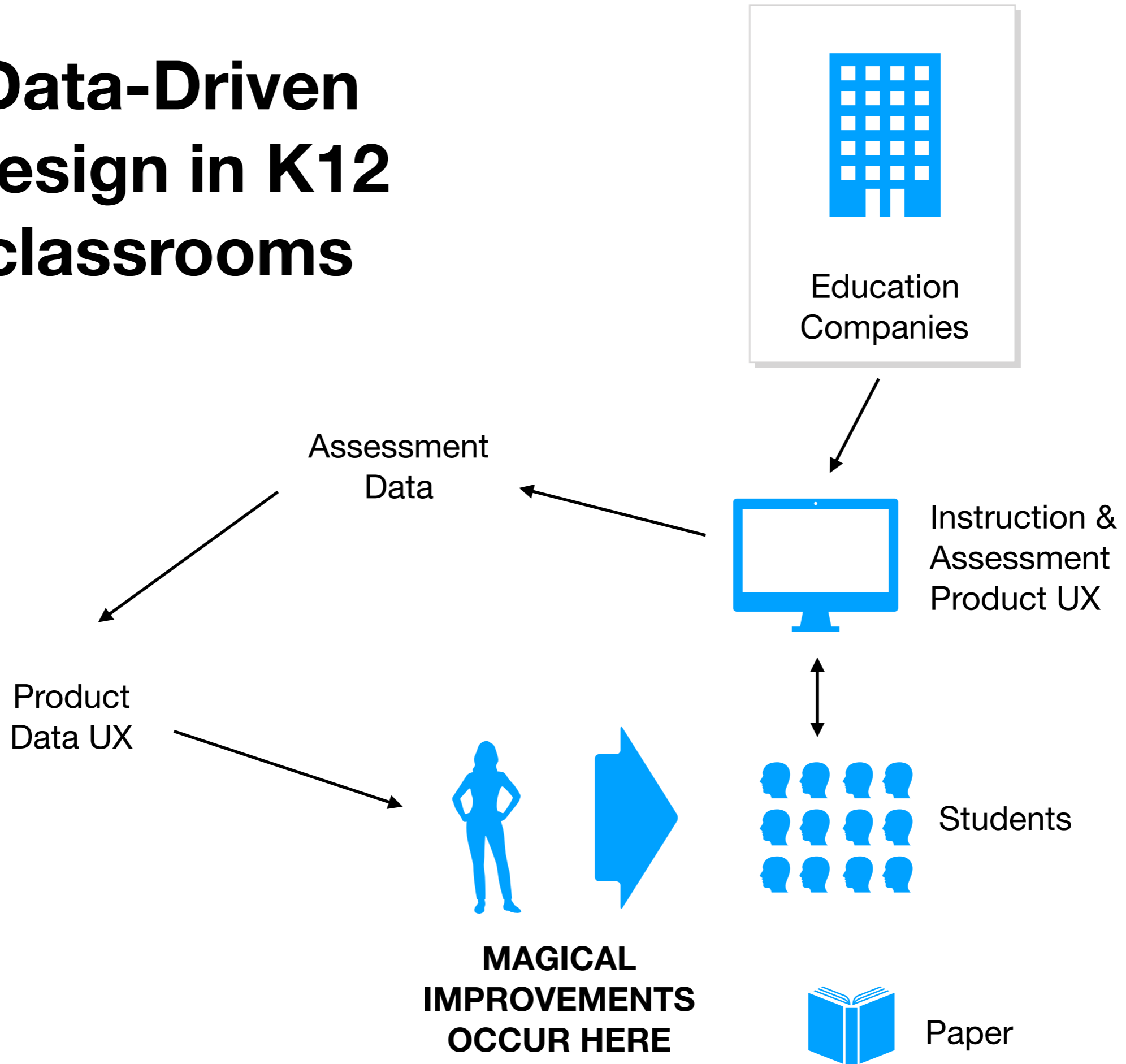
Slavin et al, 2013

1. Gather assessment data
2. Analyze data to identify problems and their causes
3. Select actions to address those problems



Slavin, R. E., Cheung, A., Holmes, G., Madden, N. A., & Chamberlain, A. (2013). Effects of a data-driven district reform model on state assessment outcomes. *American Educational Research Journal*, 50(2), 371-396.

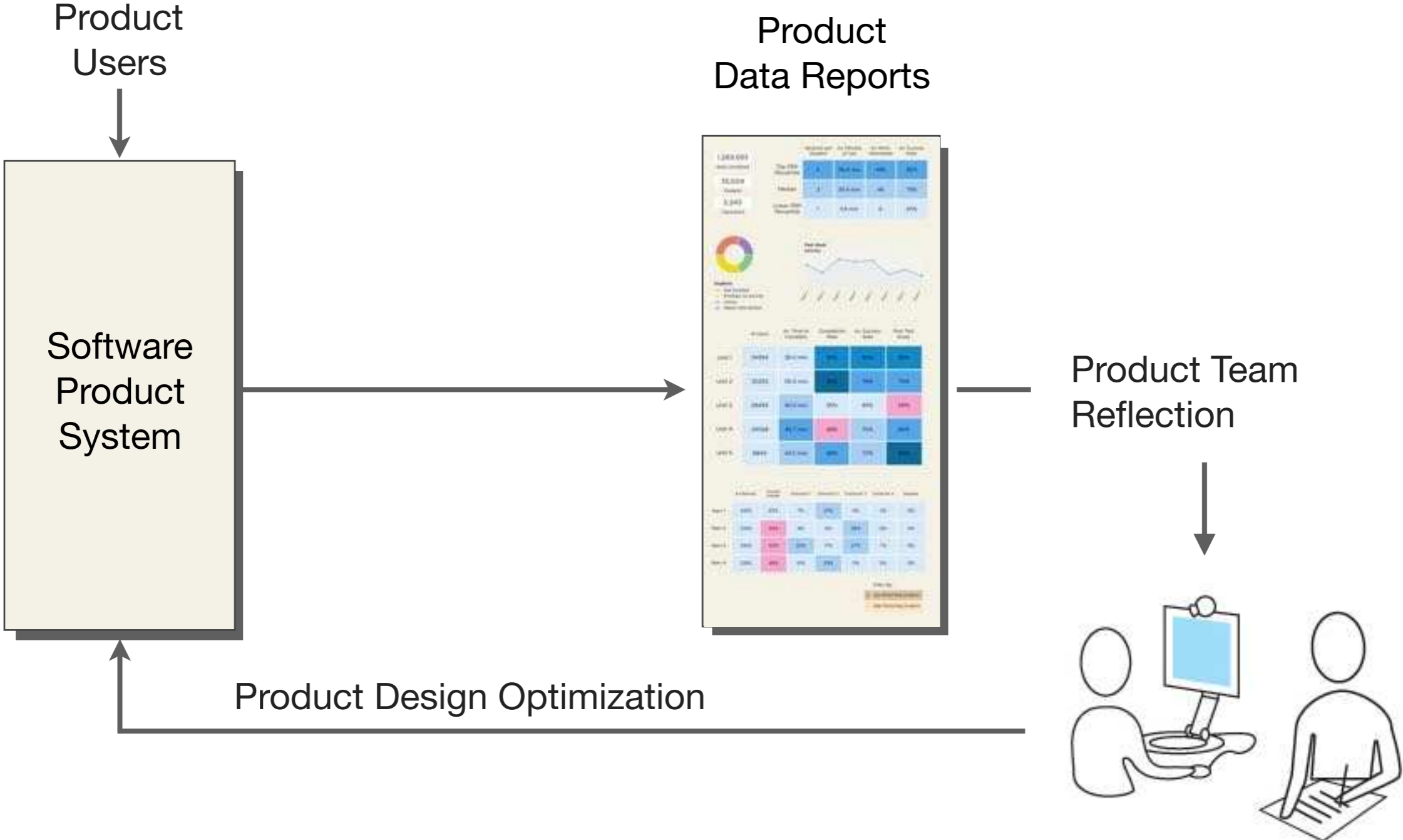
Data-Driven Design in K12 classrooms



What is Data-Driven Design?

- 1. Define “the right” goals (or needs).**
- 2. Define “the right” measures of those goals.**
- 3. Take actions to modify the current system when goals aren’t met.**

Software Model of Data-Driven Continuous Improvement



How Might We Design Systems to Support Data-Driven Continuous Improvement?

Simple Heuristic Advice:
*just make it easier to
measure and modify*

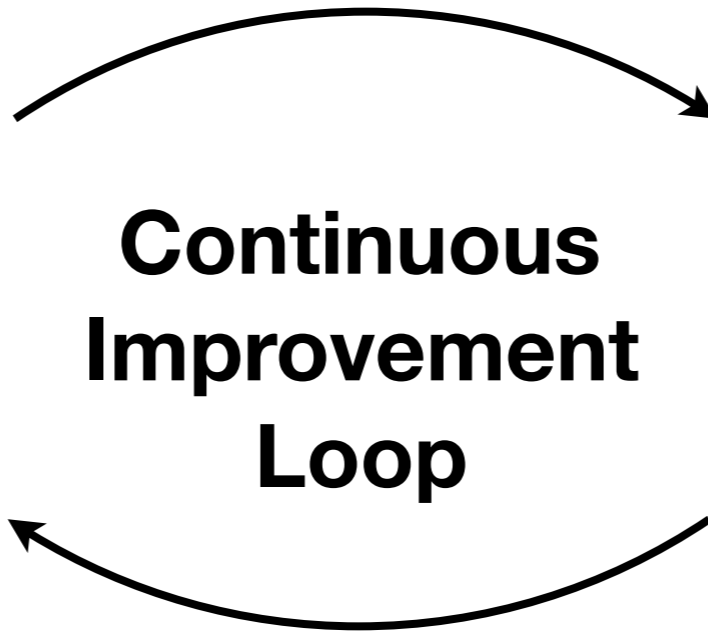
Measure
Outcomes

**Continuous
Improvement
Loop**

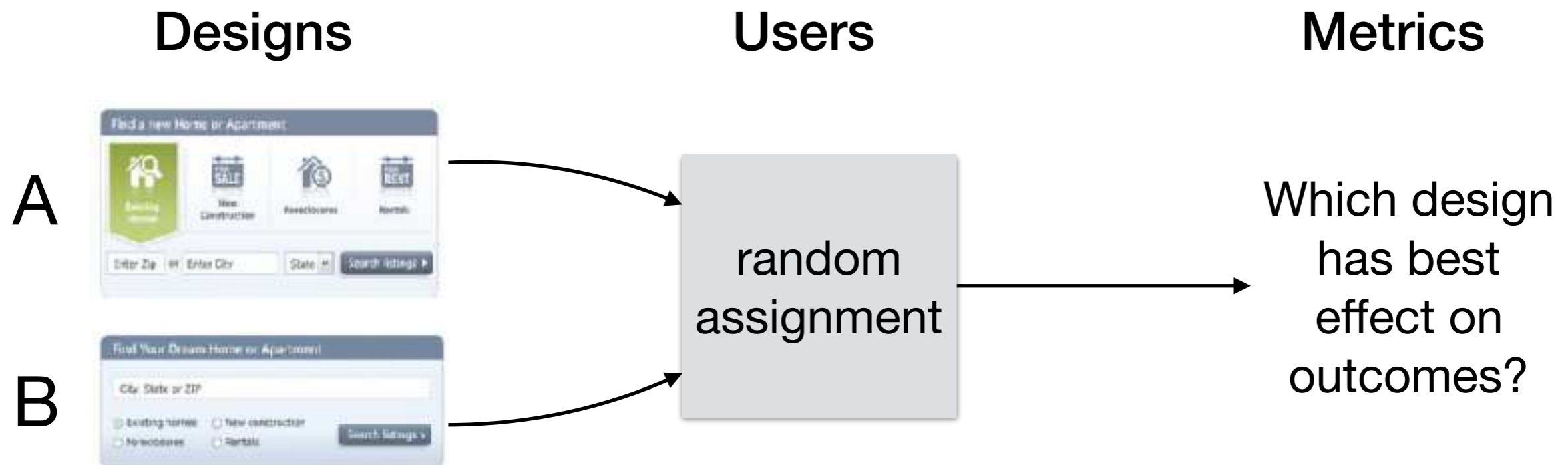
Modify
Designs

“What are the
areas of need?”

“What can we
do about it?”



Online Product Experiments (A/B Tests): A Method for Continuous Improvement



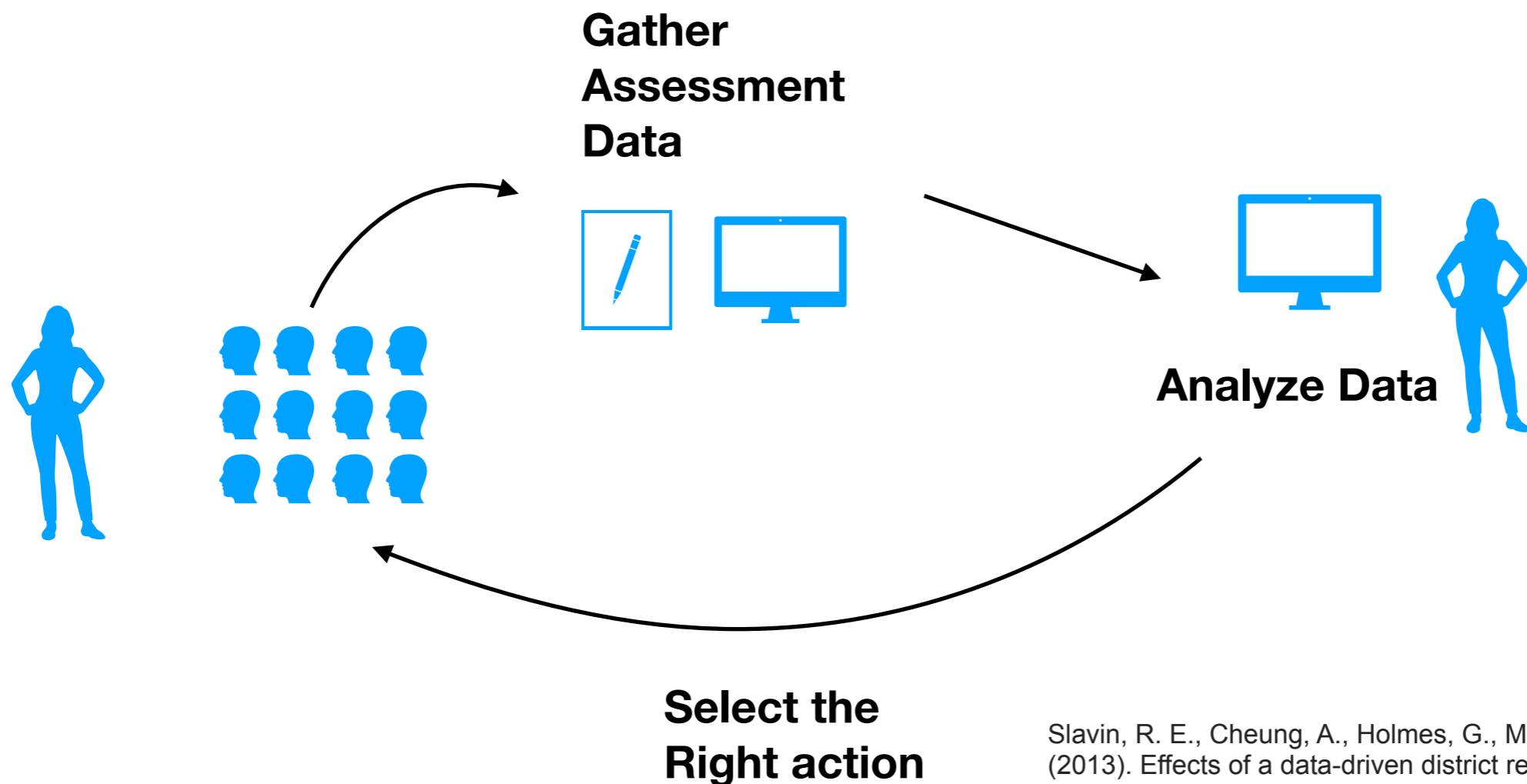
Over 10,000/day run by Google, Facebook, Amazon, etc

Continuous Improvement in K12

"Data-Driven Decision-Making"

Slavin et al, 2013

1. Gather assessment data
2. Analyze data to identify problems and their causes
3. Select actions to address those problems



Slavin, R. E., Cheung, A., Holmes, G., Madden, N. A., & Chamberlain, A. (2013). Effects of a data-driven district reform model on state assessment outcomes. *American Educational Research Journal*, 50(2), 371-396.

Why is there desire for “Data-Driven Decision Making” (DDDM) in K12?

Very few interventions have been found to be effective in addressing the poverty achievement gap.

In a review of 196 randomised field experiments, Data-Driven Decision Making was found to be one of the most effective approaches for improving high-poverty schools (Fryer, 2017).

Fryer Jr, R. G. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experimentsa. In Handbook of Economic Field Experiments(Vol. 2, pp. 95-322). North-Holland.

Van Geel, M., Keuning, T., Visscher, A. J., & Fox, J. P. (2016). Assessing the effects of a school-wide data-based decision-making intervention on student achievement growth in primary schools. *American Educational Research Journal*, 53(2), 360-394.

DDDM is not “disruptive” or radical change.

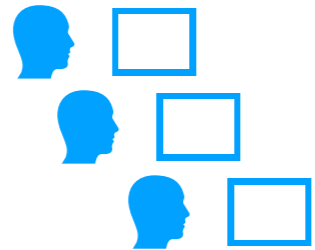
DDDM is all about small incremental improvements... ideal for complex social-technical systems

**K12 Education is a
Complex-Technical System**

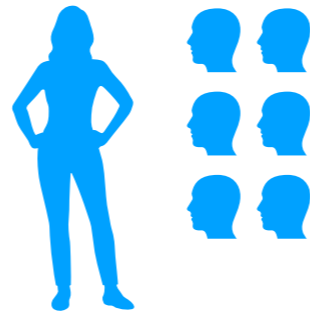
K12 Education is a Complex-Technical System



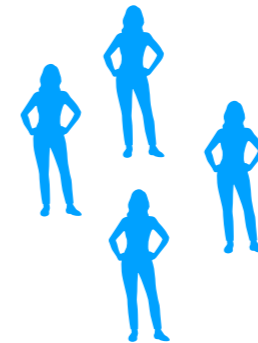
Family Interactions



Computer Interactions



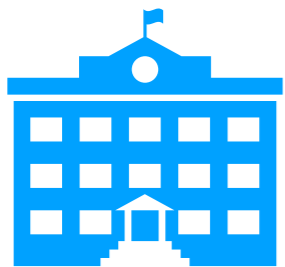
Teacher Interactions



PLCs



Government Policy



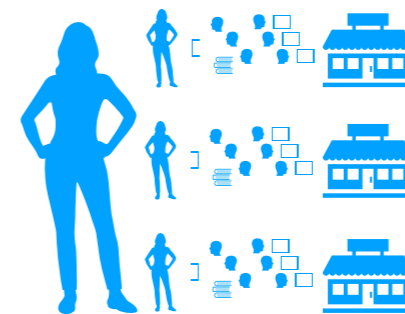
Out-of-School Programs



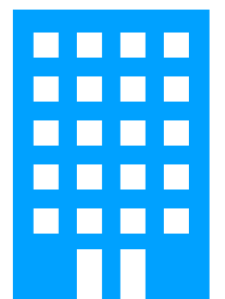
Text Interactions



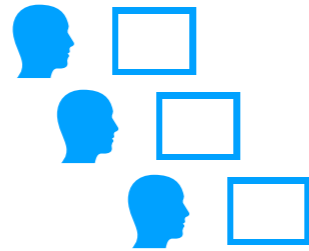
Small-Group Interactions



Administrative Interactions



Education Companies



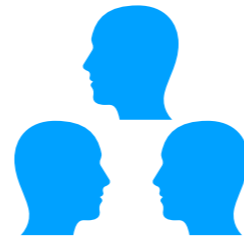
Computer
Interactions



Teacher
Interactions

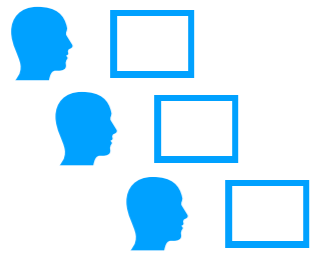


Text
Interactions

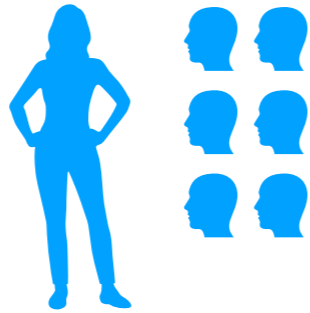


Small-Group
Interactions

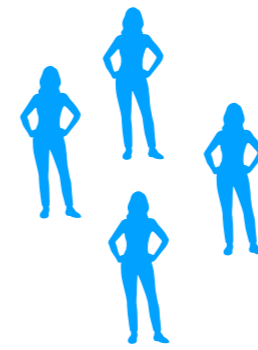
The Modern Classroom



Computer Interactions

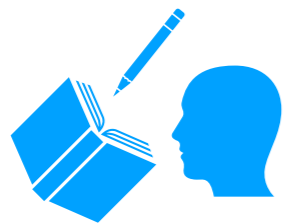


Teacher Interactions

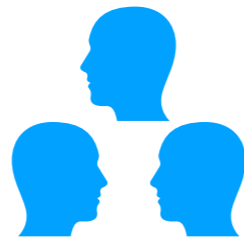


PLCs

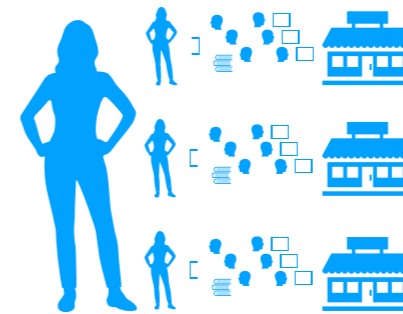
“Professional Learning Communities”



Text Interactions

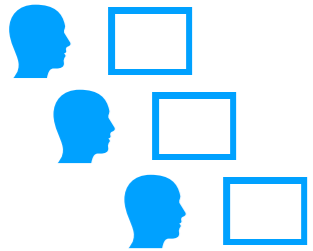


Small-Group Interactions

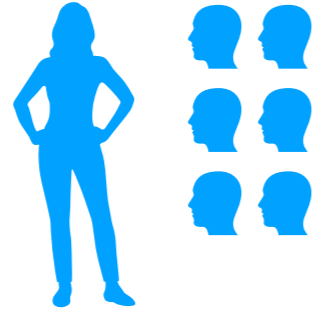


Administrative Interactions

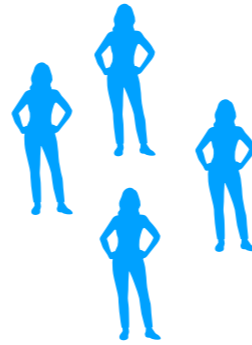
- Curriculum Coordinators
- Schools
- Districts
- State Admin



Computer Interactions



Teacher Interactions

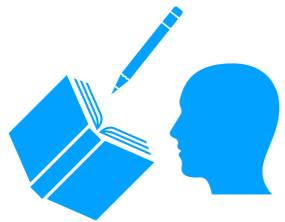


PLCs



Government Policy

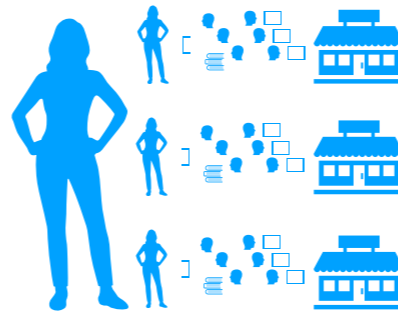
- Learning Standards



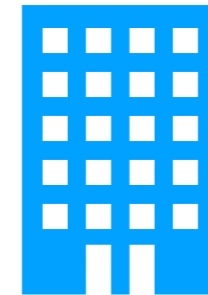
Text Interactions



Small-Group Interactions



Administrative Interactions



Education Companies

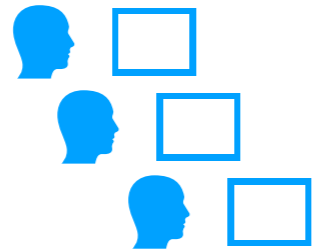
- Textbooks
- Assessments
- Paper & Digital Products

K12 Education is a Complex-Technical System

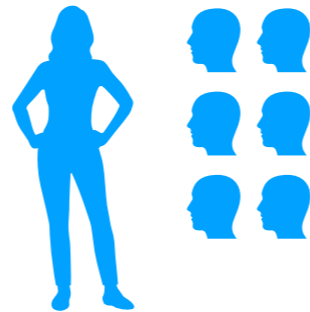
Opportunities from Systems Design?



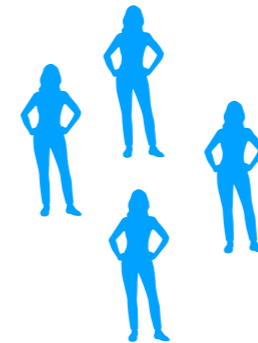
Family
Interactions



Computer
Interactions



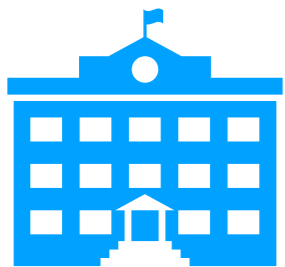
Teacher
Interactions



PLCs



Government
Policy



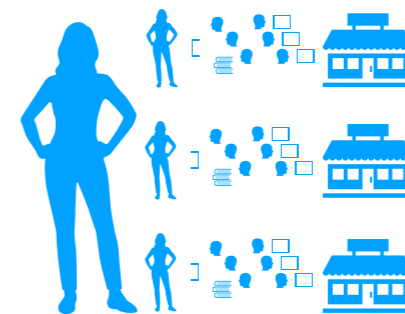
Out-of-School
Programs



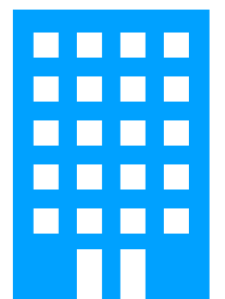
Text
Interactions



Small-Group
Interactions



Administrative
Interactions



Education
Companies

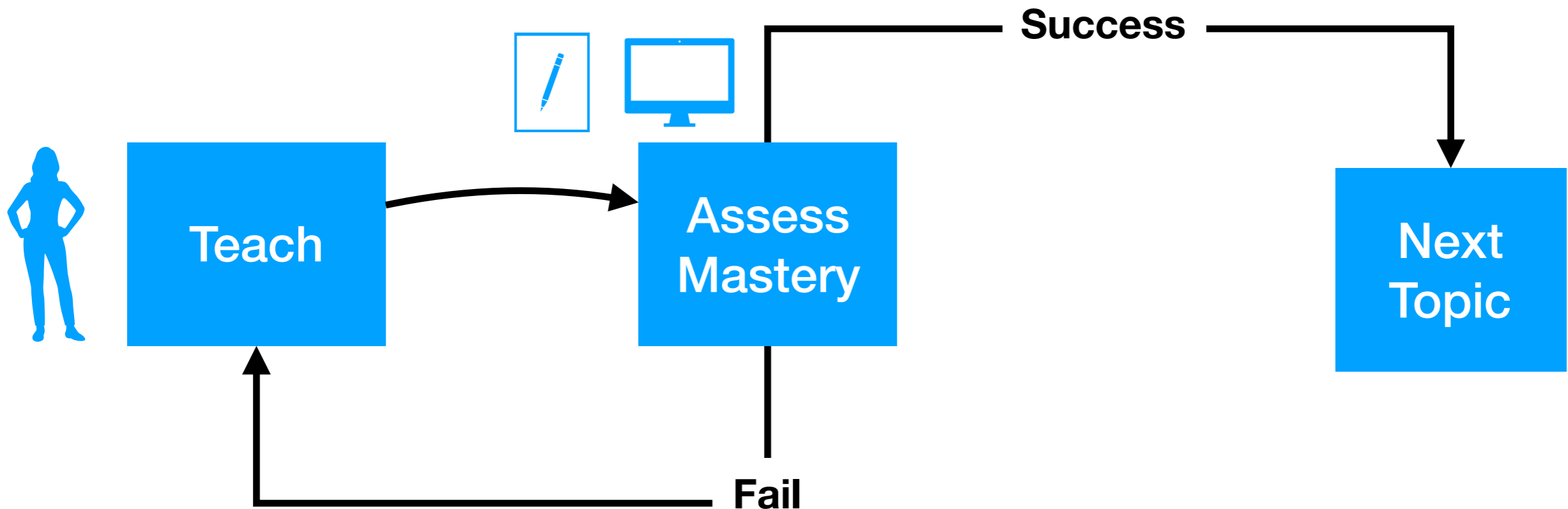
Systems Design

Role of data in systems design

- Systems use data as feedback to guide action.
- Measurements provide data to inform systems whether goals have been achieved.
- Cybernetic systems?

Mastery Learning

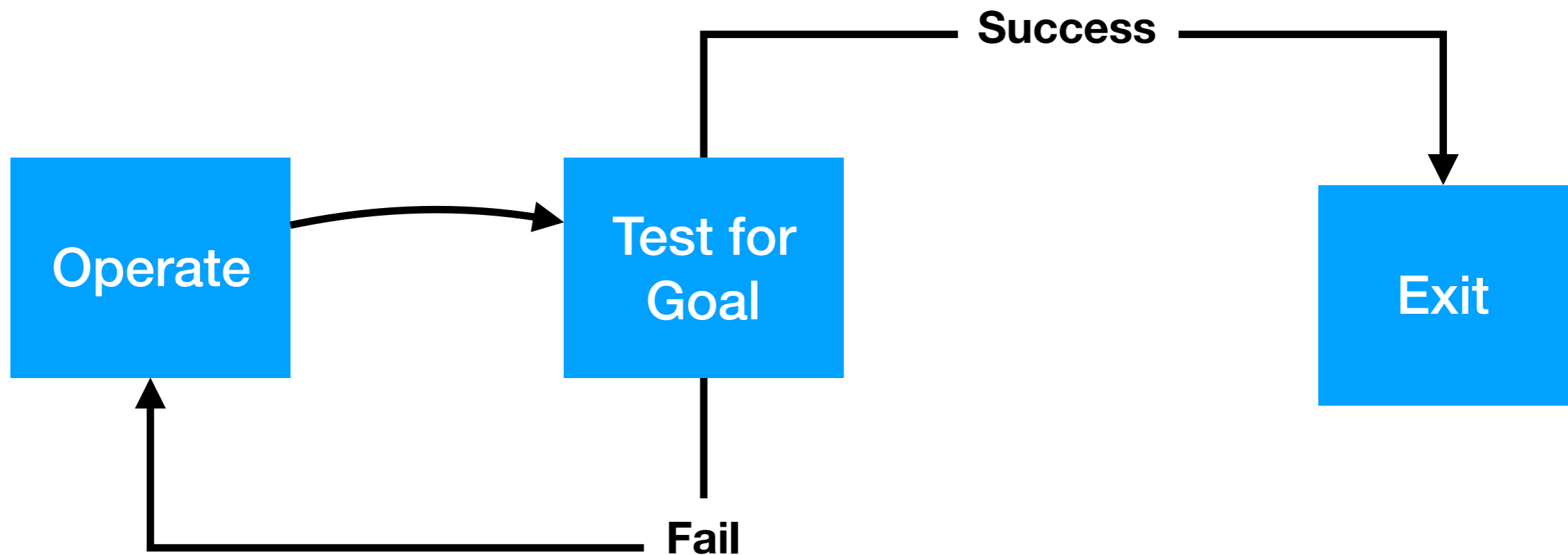
as a cybernetic system



Closed System

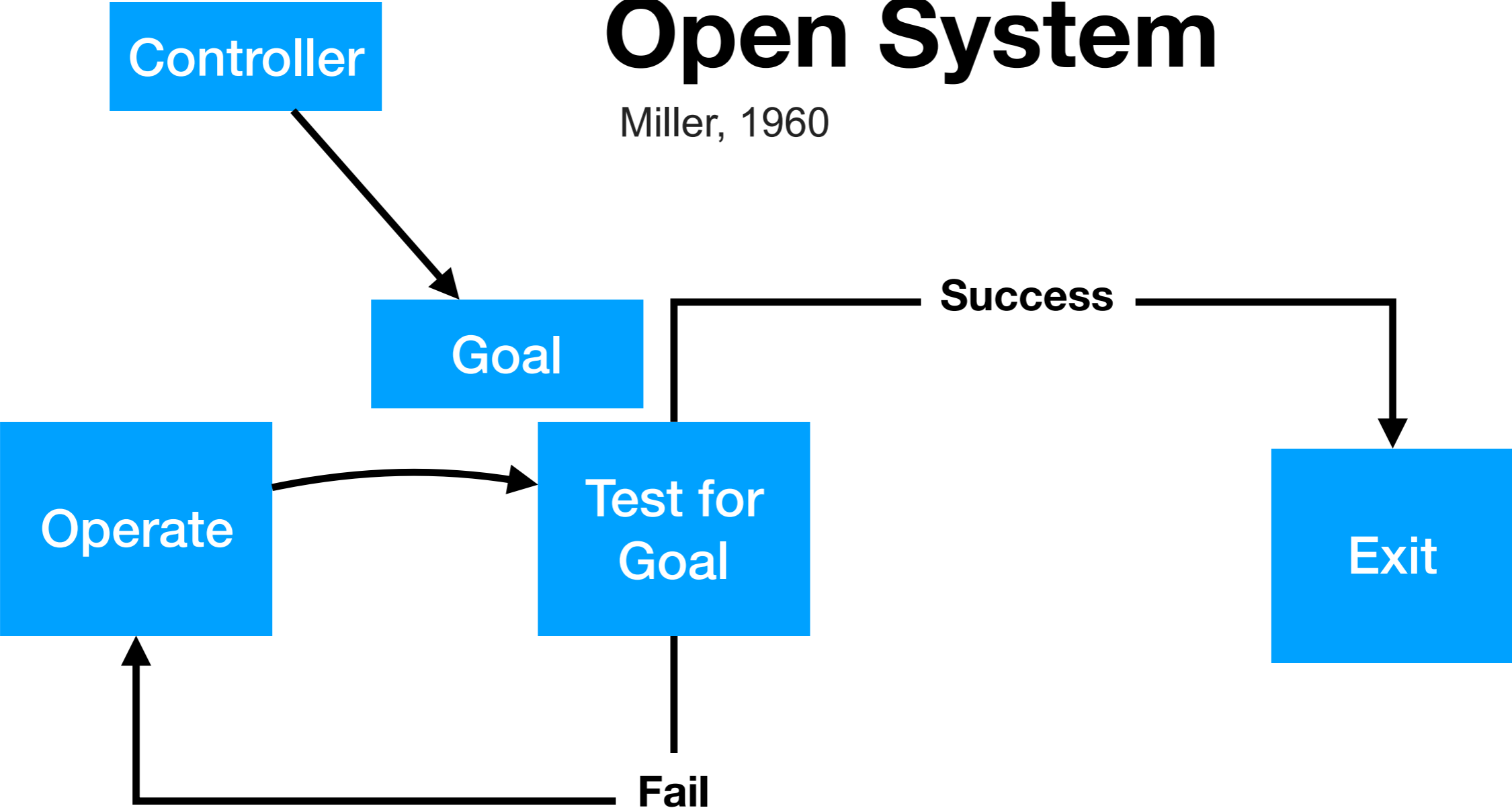
TEST, OPERATE, TEST, EXIT

Miller et al, 1960



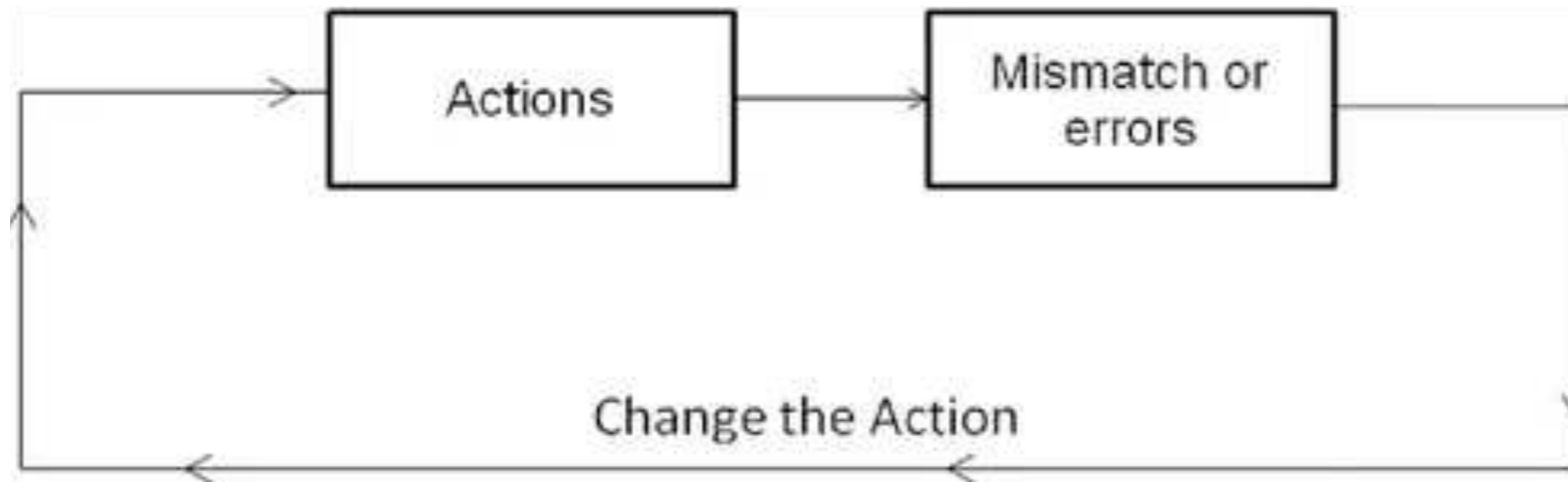
Open System

Miller, 1960

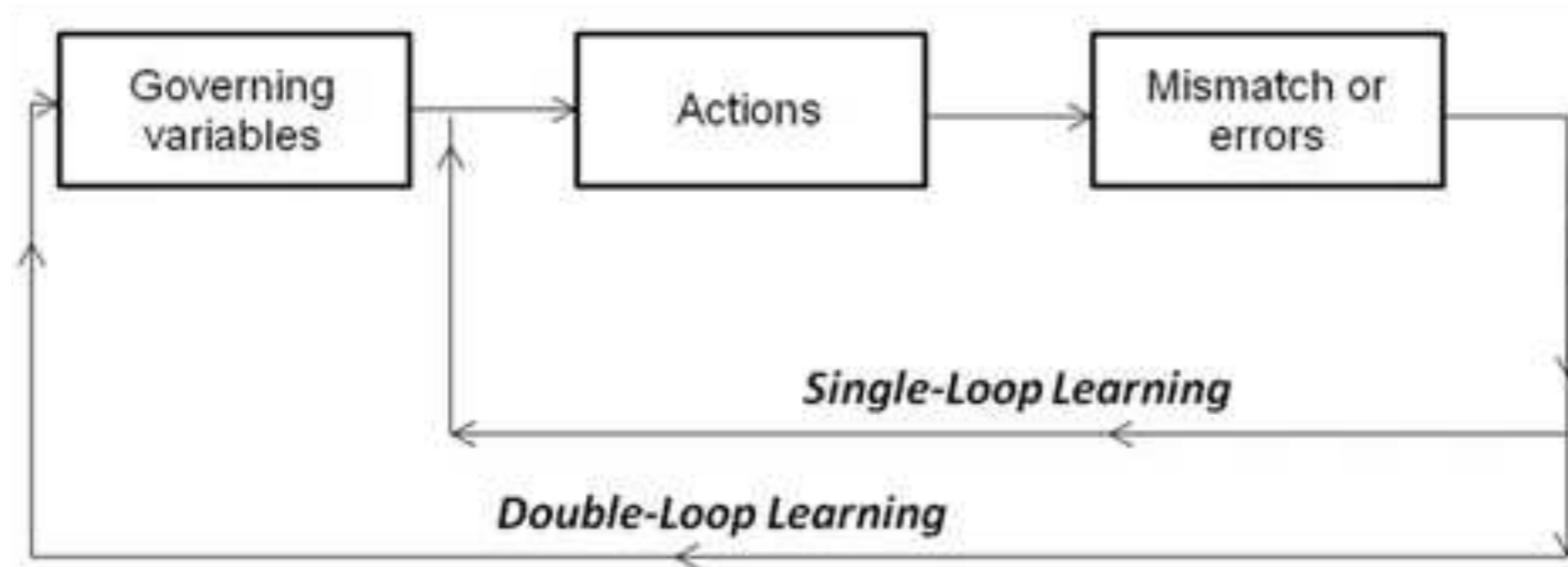


Single and Double Loop Learning

Argyris 1990

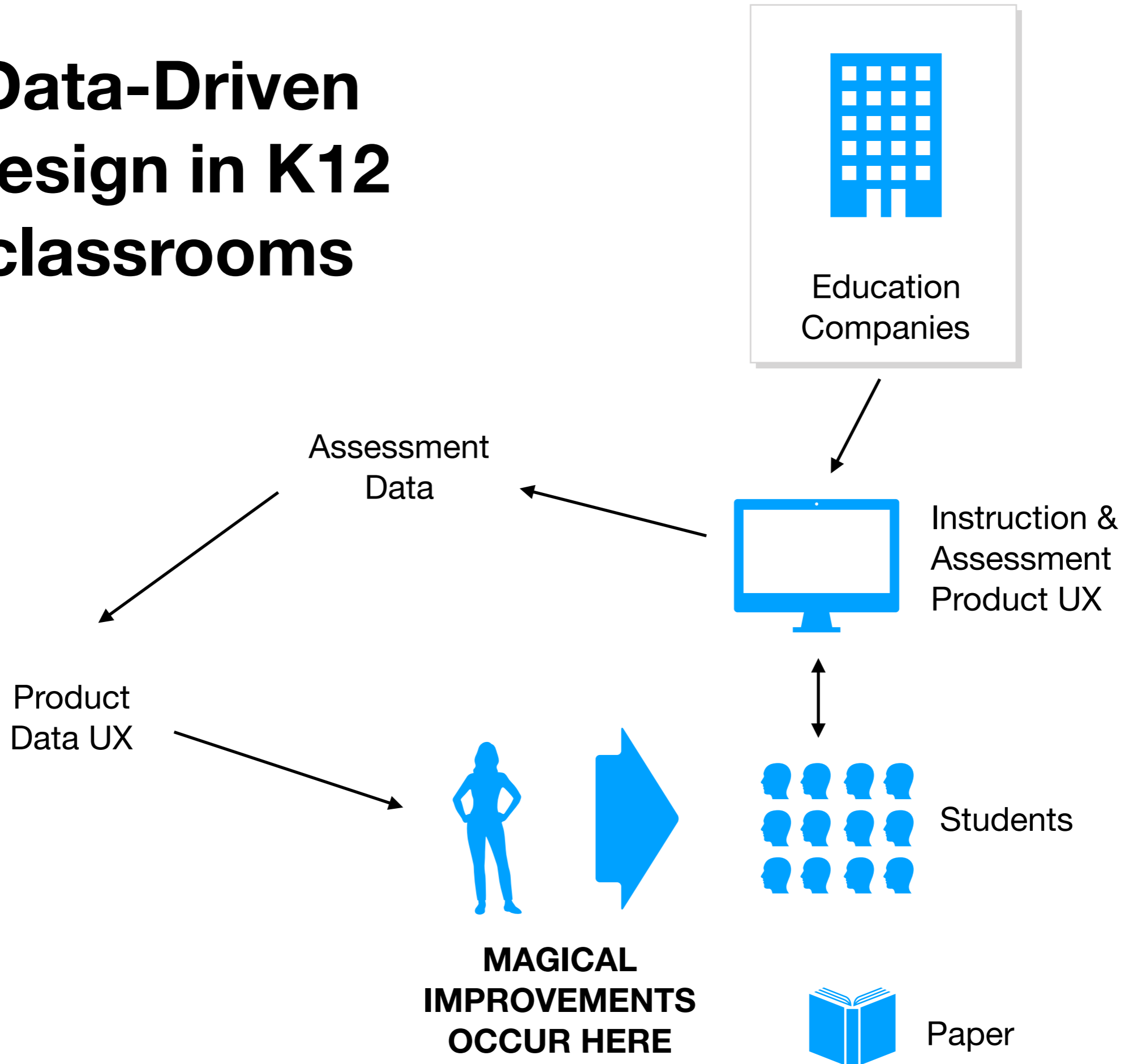


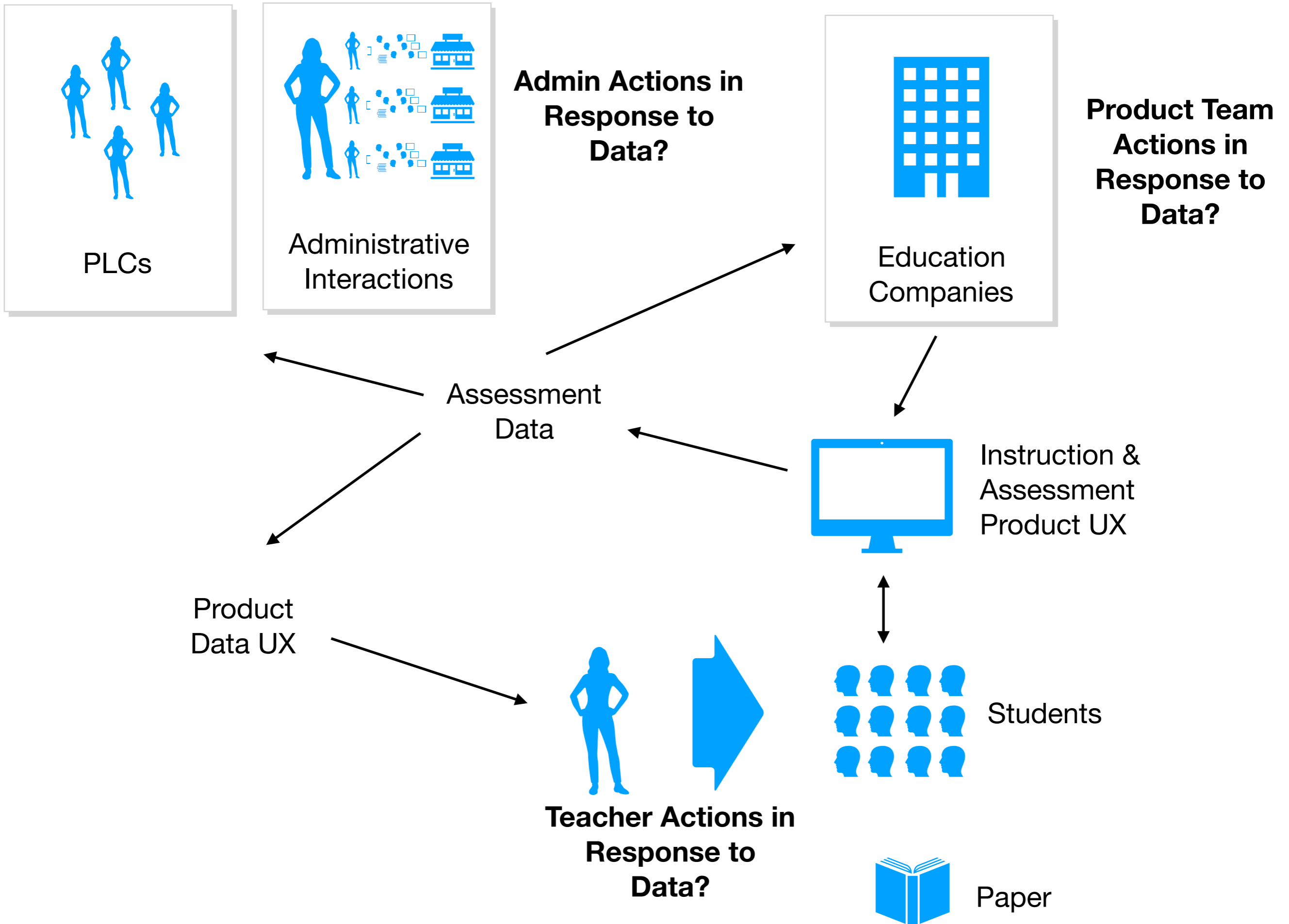
p92



p94

Data-Driven Design in K12 classrooms



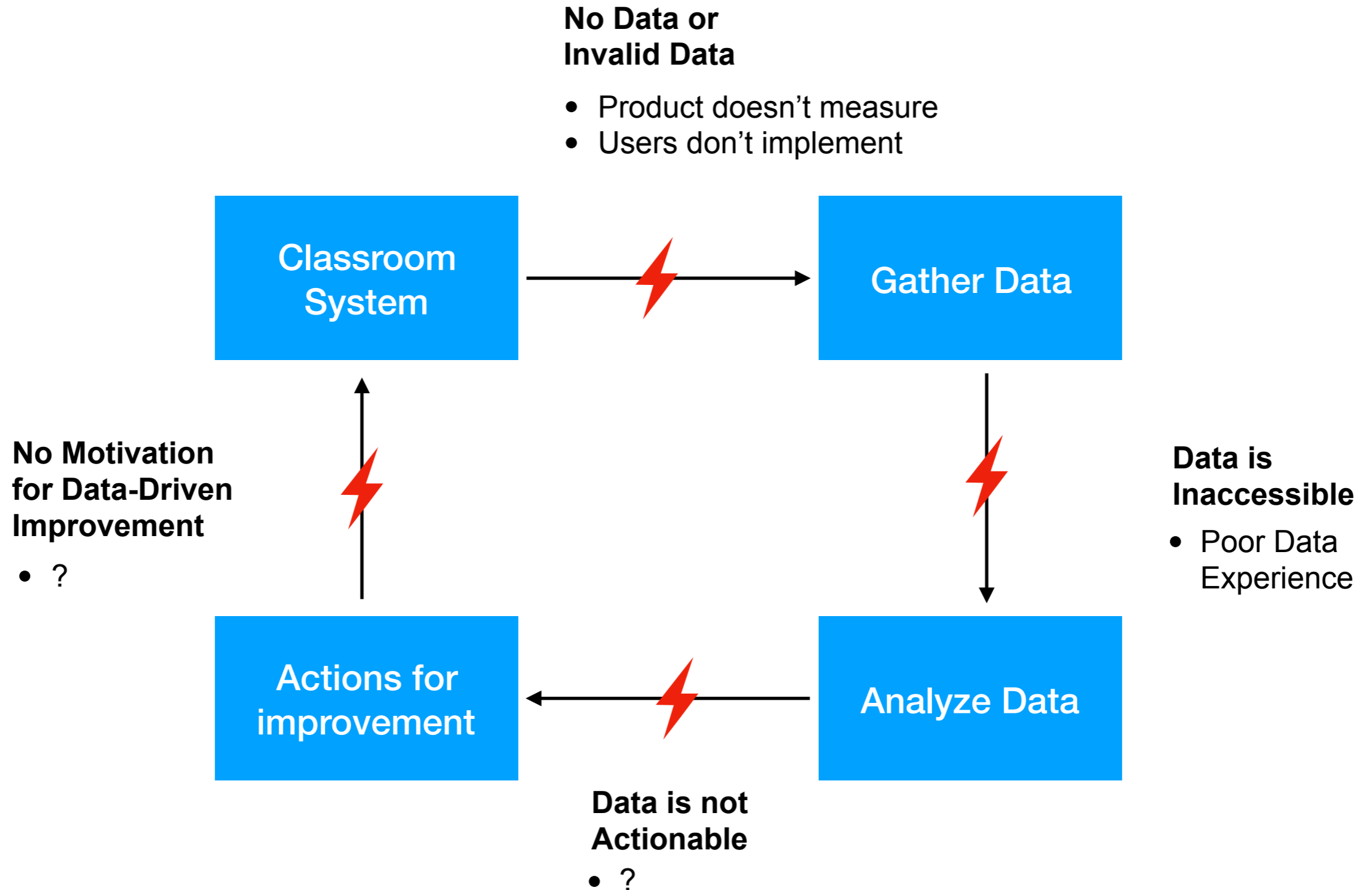


Case #1:

**Supporting Data-Driven
Improvement in Educational
Product Companies**

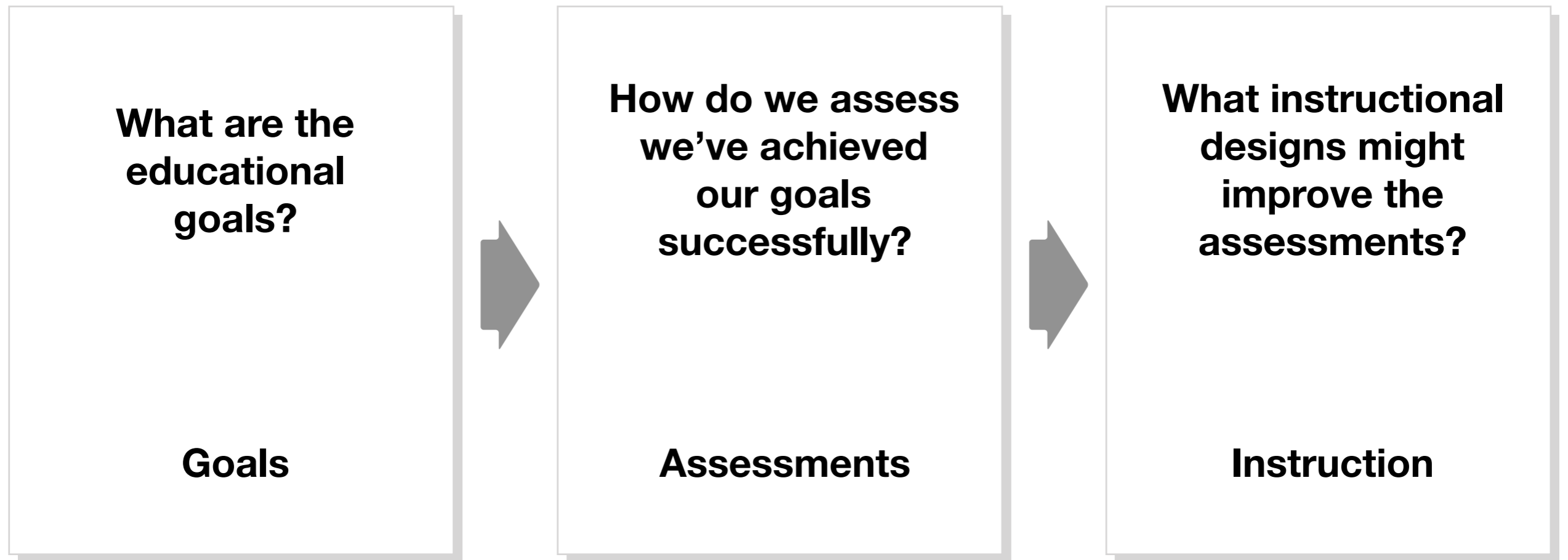
Common Barriers to DDDM

And some root causes



"Backwards Design" Methodology

(cite Understanding by Design)



Key Question #1:

Is the system capable of measuring success?

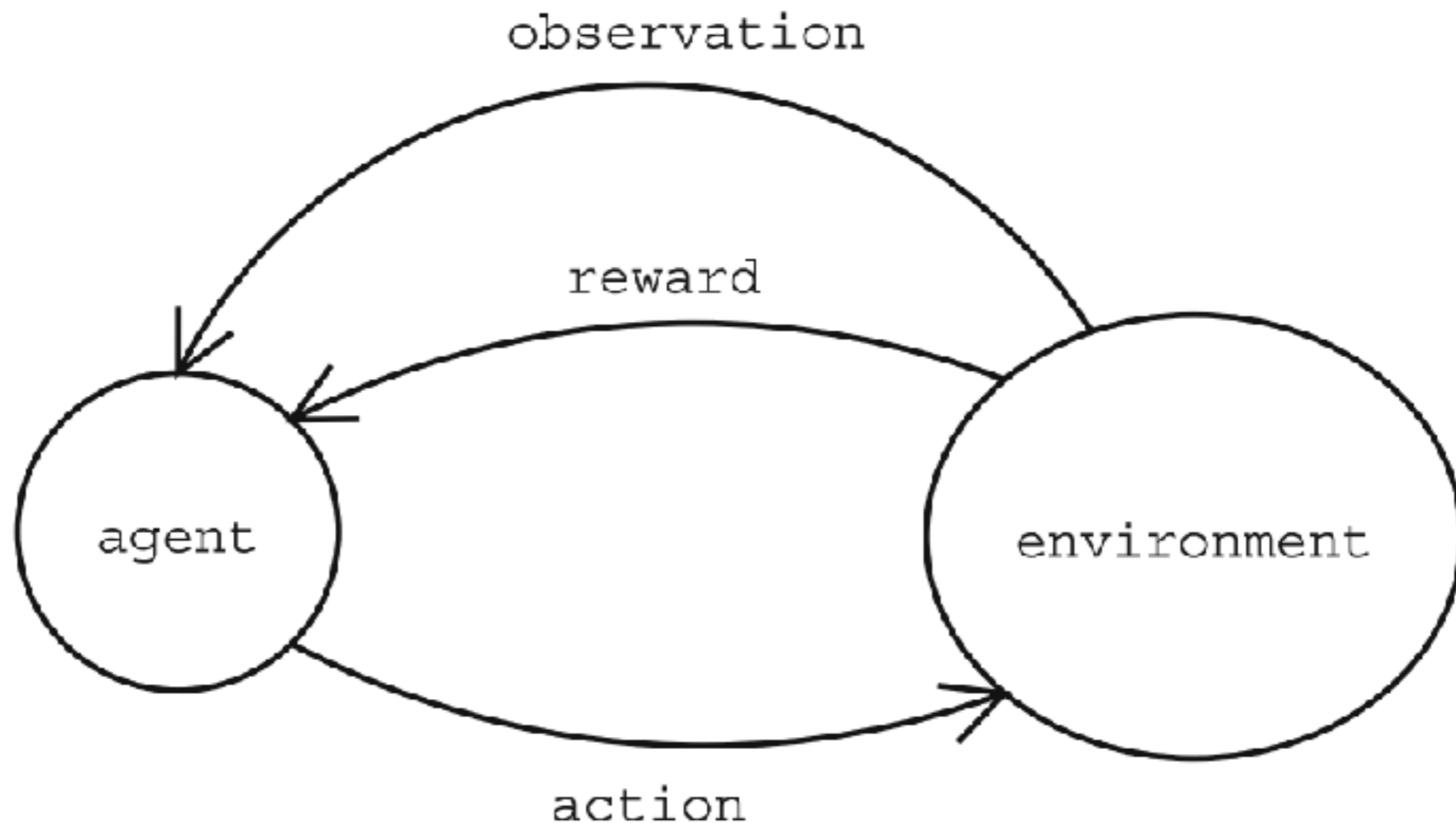
Can the system measure whether it is achieving its goals?

“If we want to report on student mastery for topic A, we need to collect valid test data on topic A”

Case #2:

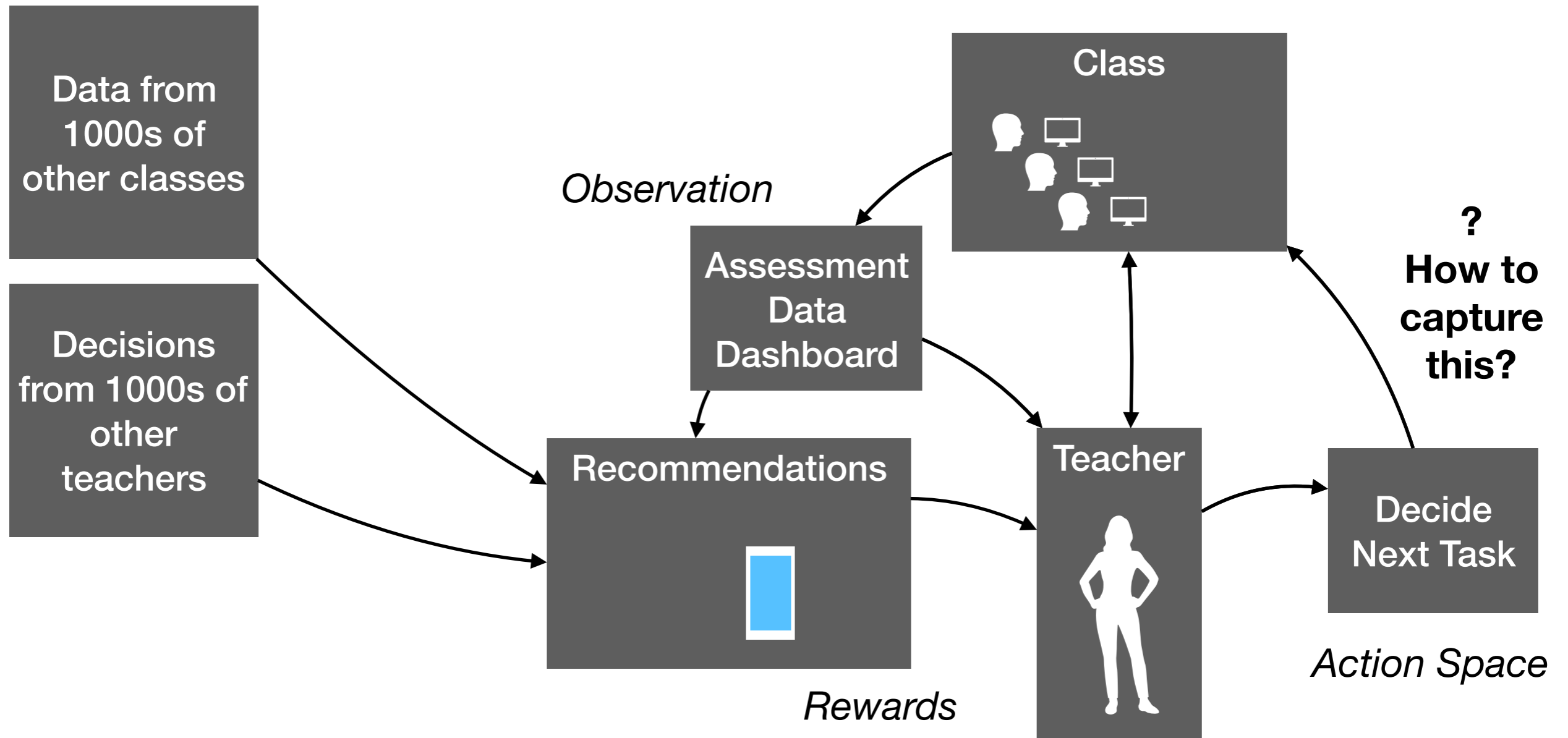
**Development plans for a
teacher-facing
recommendation system**

Mapping the System to Reinforcement Learning



- **What is the space of possible observations? (what we measure?)**
- **What is the space of possible actions? (what can we do?)**
- **What is the reward metric? (how do we quantify our goals?)**

Mapping Classrooms to Reinforcement Learning



1. Assessment Data Dashboard (*Observation*)
2. Digital Assignment Possibilities (*Action Space*)
3. Recommendation Success (*Rewards*)

Key Question #2:

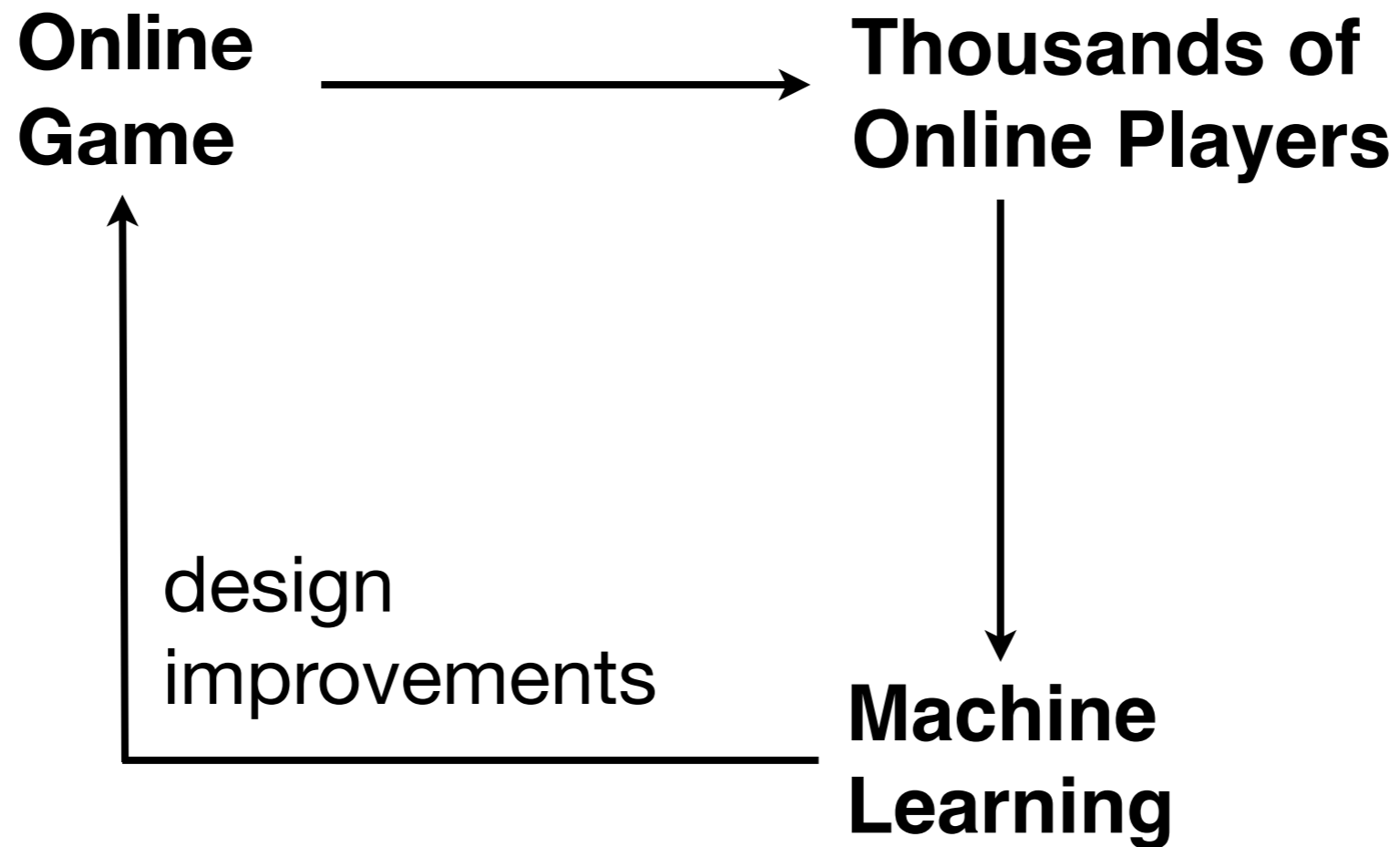
Is the action space sufficiently defined?

The right data might be gathered and analysed appropriately ...but is there a defined action in response?

NEED methods for scaling-up system learning from rich human decision-making

Case #3:

**Automated Machine-Learning
Optimization of an Educational
Game**



An AI system for design optimization successfully increased game engagement...

...but then spun out of control!

Lomas, D., Forlizzi, J., Poonwala, N., Patel, N., Shodhan, S., Patel, K., Koedinger, K., Brunskill, E. (2016) Interaction Design as a Multi-Armed Bandit Problem. *ACM CHI*

Multi-Armed Bandit Problem

Goal is to maximize payout from row of slot machines with unknown rates of reward (some machines pay out more)

Need to balance *exploration* with *exploitation*



Interface design as a multi-armed bandit problem...

Lomas, J. D., Forlizzi, J., Poonwala, N., Patel, N., Shodhan, S., Patel, K., Koedinger, Brunskill, E. (2016).
Interface Design Optimization as a Multi-Armed Bandit Problem. In CHI'16.

Instructional Design of “Battleship Numberline”



Ship targets

Type to name a fraction



Submarine targets

Click to locate a fraction

Design Factors

Target Type: Ship or Sub

Target Size: Small to large targets

Time Limits: Small to large time limits

Item Sets: The fractions to be estimated

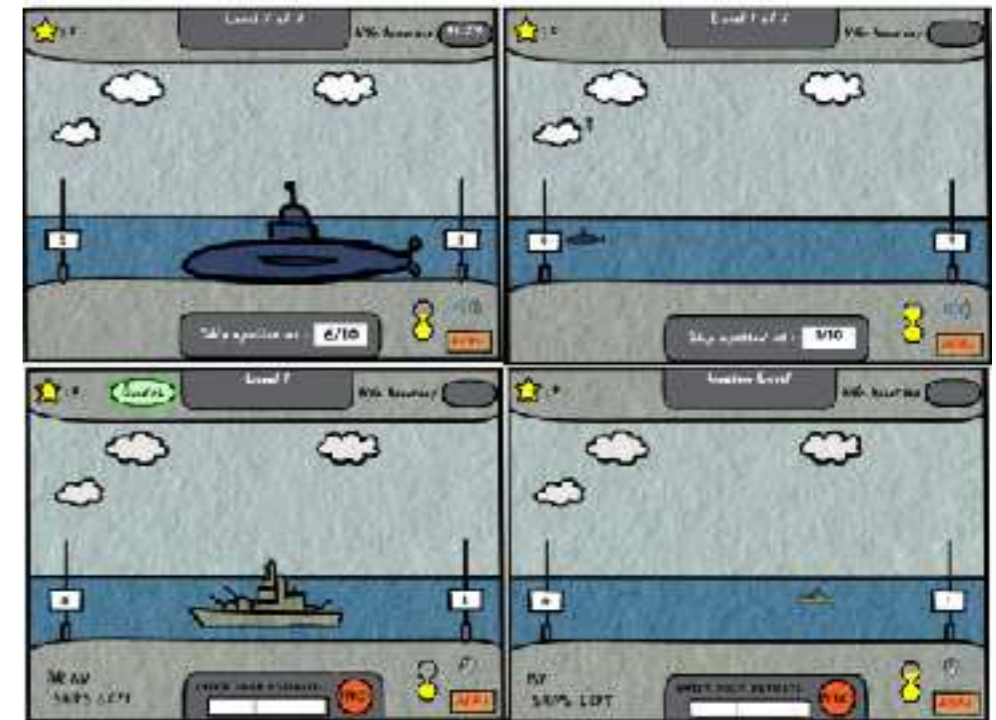
Tickmarks: Scaffolds for estimating

Sequencing: Algorithms for presentation

Experiment 1: Assigned 10,832 players to 3 simultaneous experiments, each testing 6 different design factors (2x3: target type and target size)

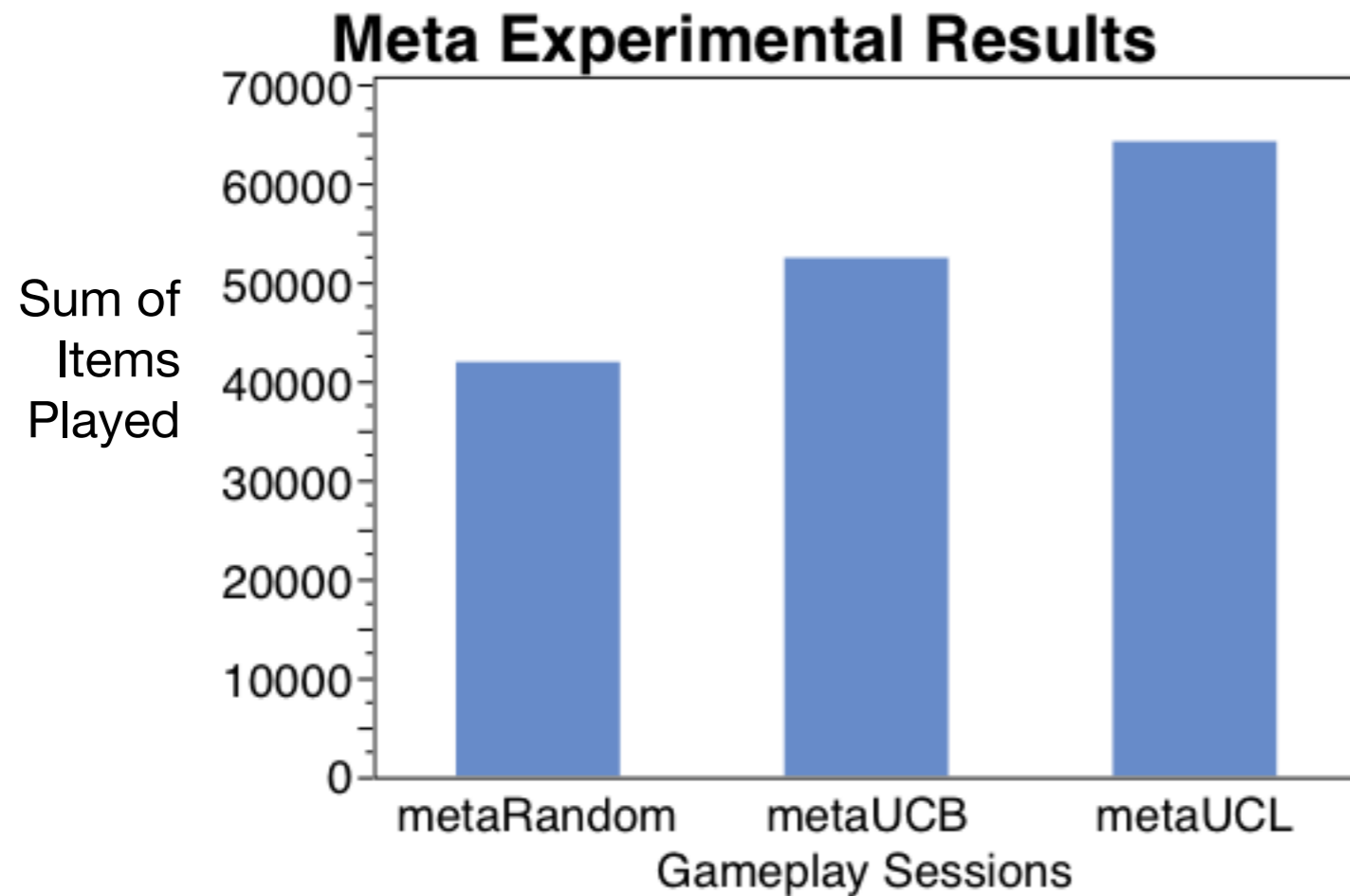
Target Type

Target Size



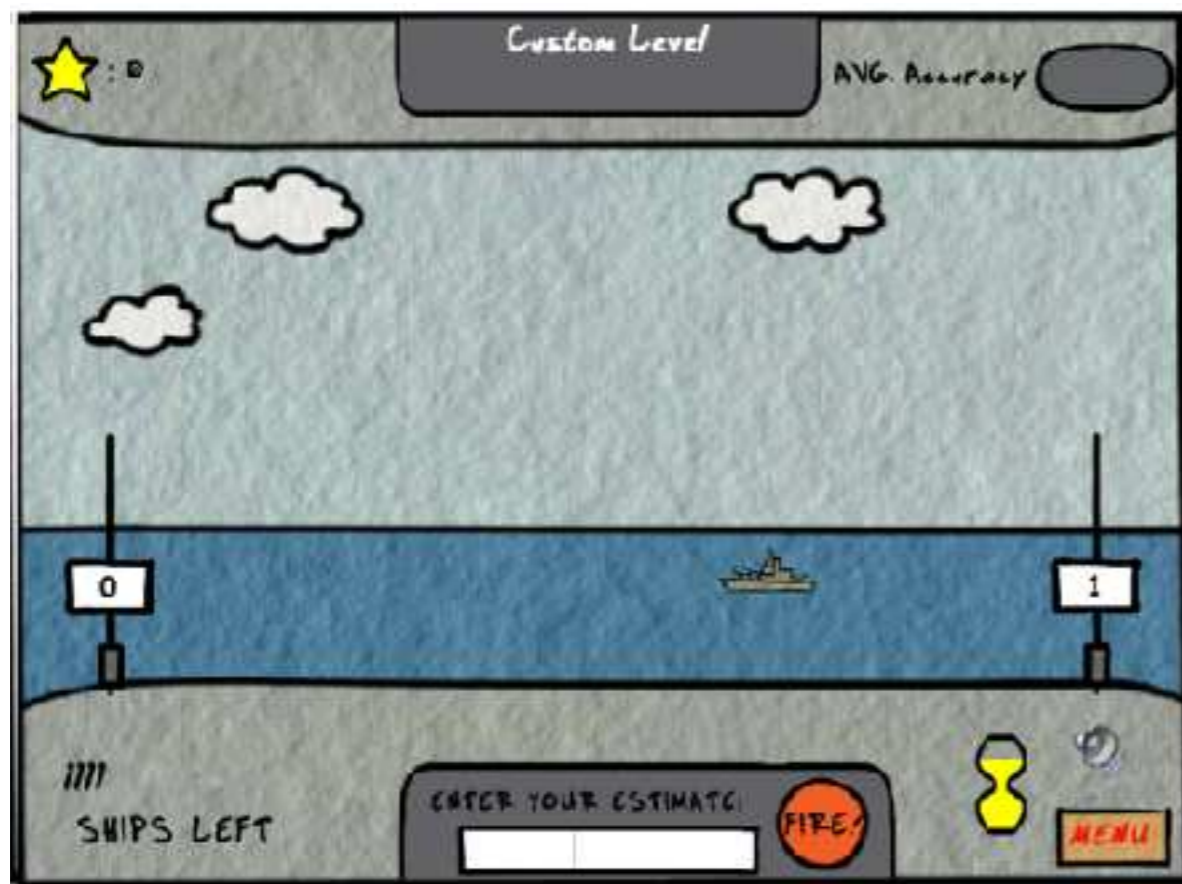
Success!

Automatic optimization algorithms produced more engagement

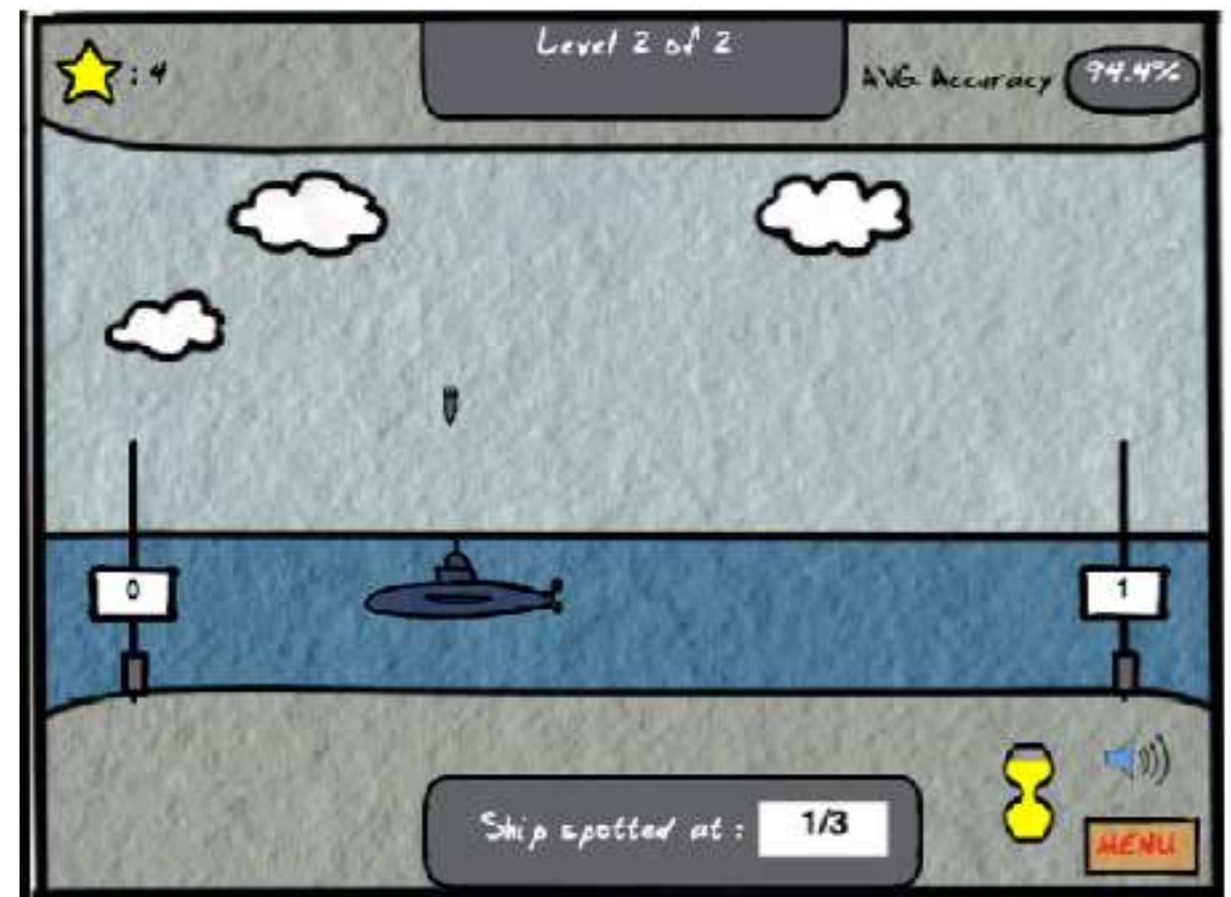


Bandit experiments
produce greater
total engagement

Meta Experiment 1



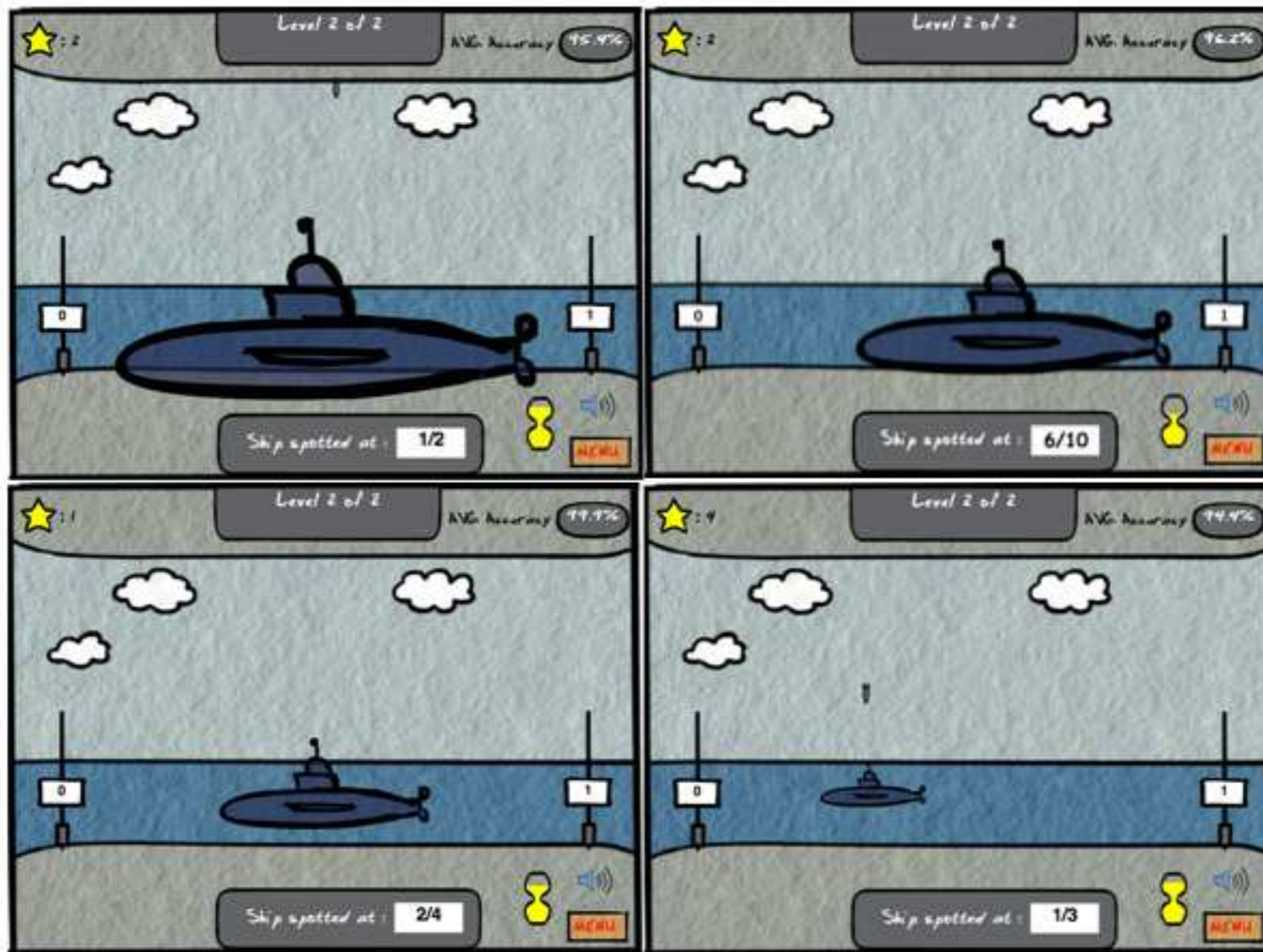
Worst Design
Ship 97

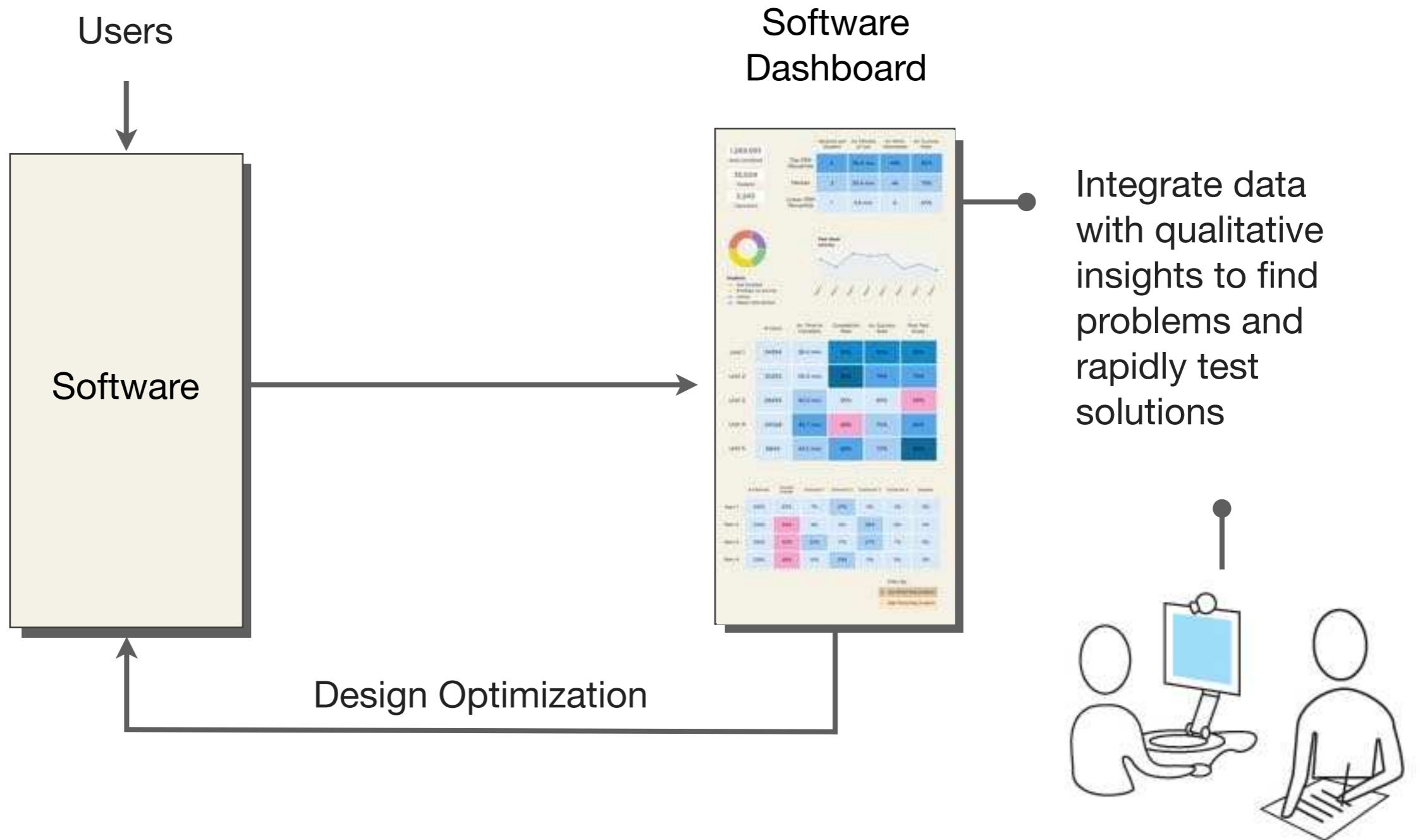


Best Design
Ship 90

Meta Experiment 2

Introduced ridiculously large "Bad Designs"





The Data-Experience Dialectic
 Data and qualitative insights must be integrated

With 10,000 subjects a day, we could run thousands of experiments each year

But setting up, analyzing and acting upon experiments is hard.

How might AI get involved? Could we partially automate design optimization and scientific experimentation?

Key Question #3:

Is there a human-in-the-loop to keep alignment?

If continuous improvement metrics are not aligned to goals and values, data-driven improvement (teams or AI) will produce unintended consequences

NEED feedback loops about suitability of success metrics

Conclusions

Takeaways

- **From Designing Artificial Intelligence to Designing Intelligent Systems**
 - **Collaboration between teams and algorithms**

(not just for AI but for products)

A Definition of Intelligence used in AI

“Intelligence measures an agent’s ability to achieve goals in a wide range of environments.”

—Legg & Hutter, 2007

“an individual’s intelligence is related to their ability to succeed...”

- Robert Sternberg

What can go wrong?

- Poor goals (misalignment with values)
- Poor metrics (misalignment with goals)
- Poor actions responding to data
- *Misleading* data and *misreading* data
- Conflicting interests in stakeholders

Unintended consequences!

Accountability can create perverse incentives

e.g. schools encourage low-performing students to drop out or to cheat (Schildkamp et al, 2012).

How can system designers help build the human into the loop?

- **UI for User Goal Setting**
- **System mapping & communicating future systems**
- **Processes to Negotiate Metrics for Success**

(not just for AI but for products)

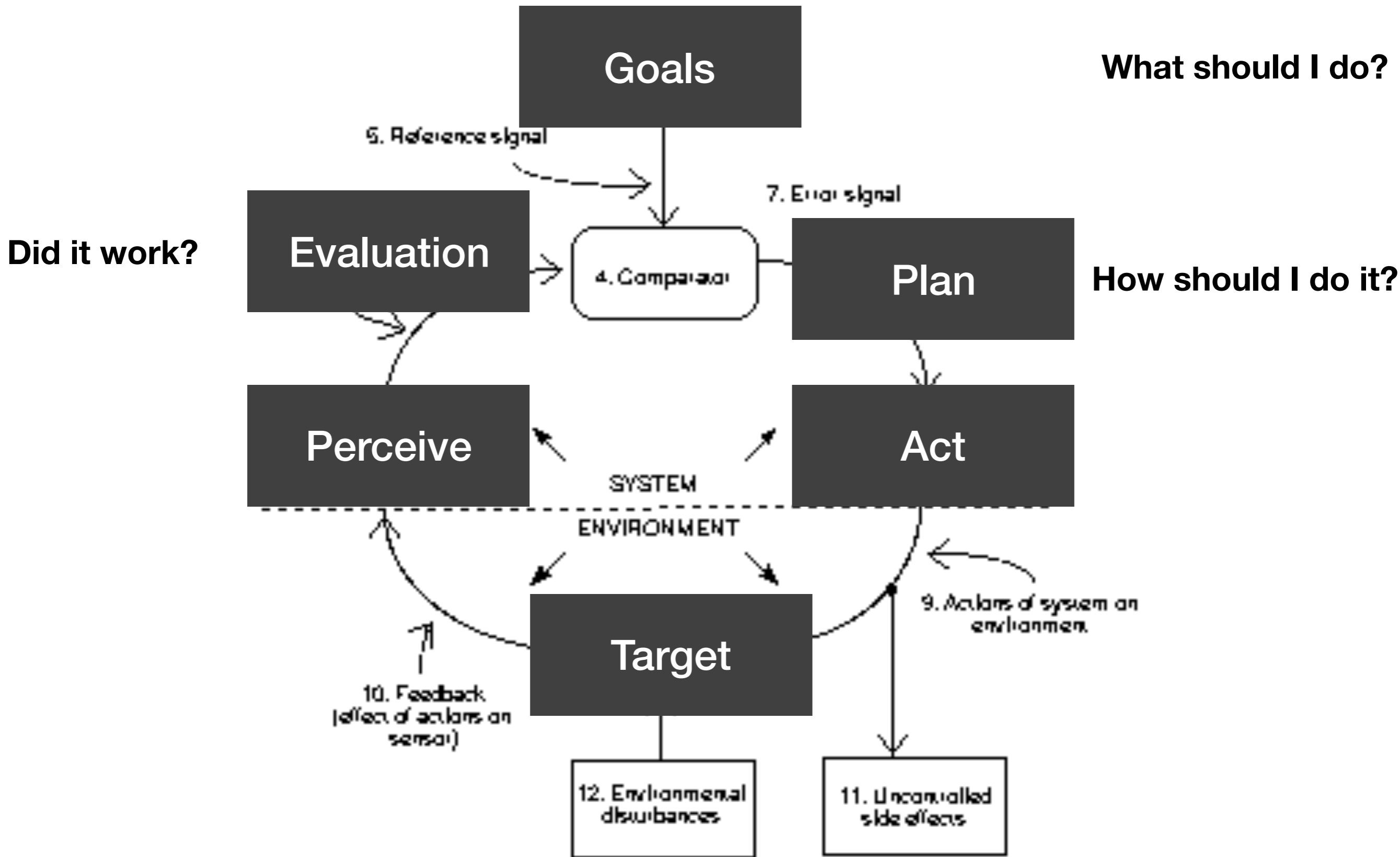
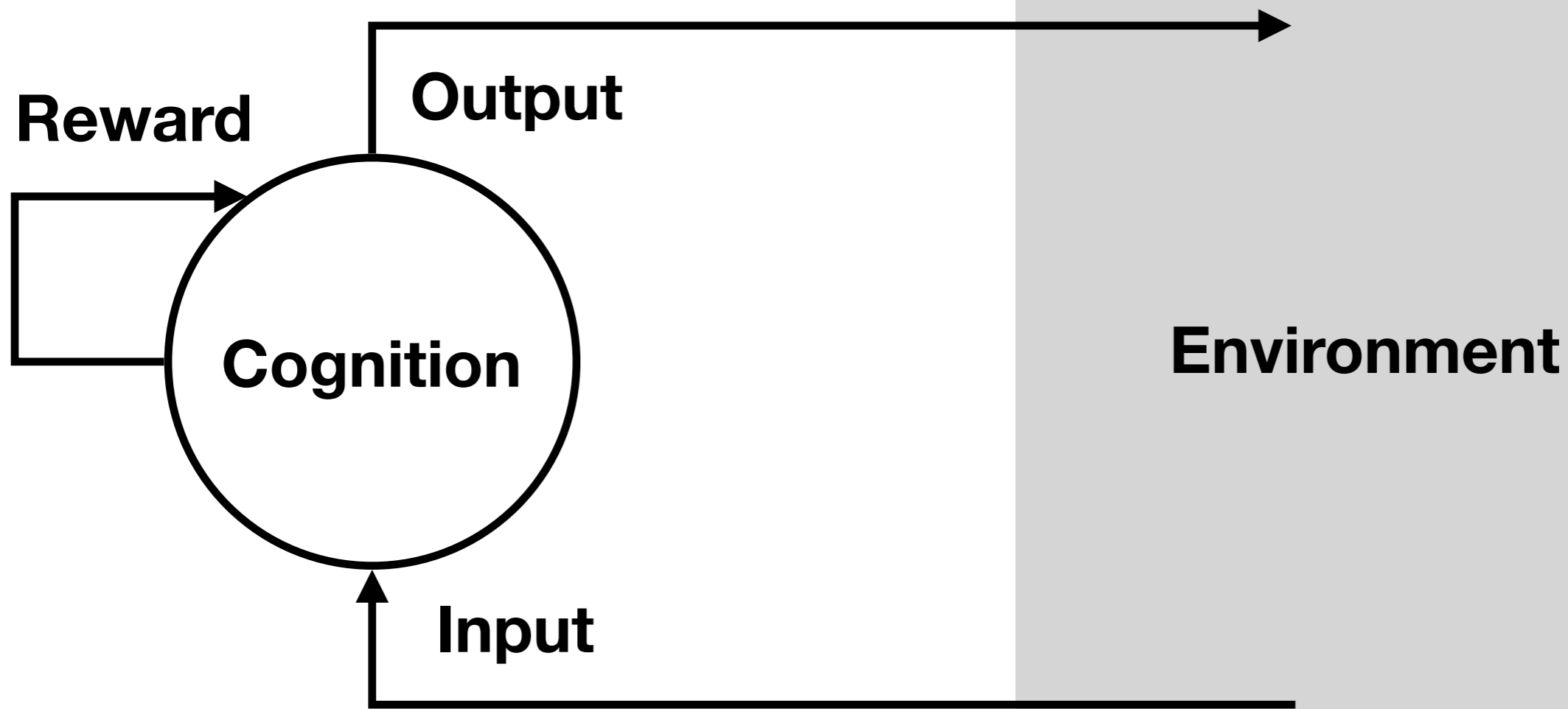


Figure 6.2
Elementary control system

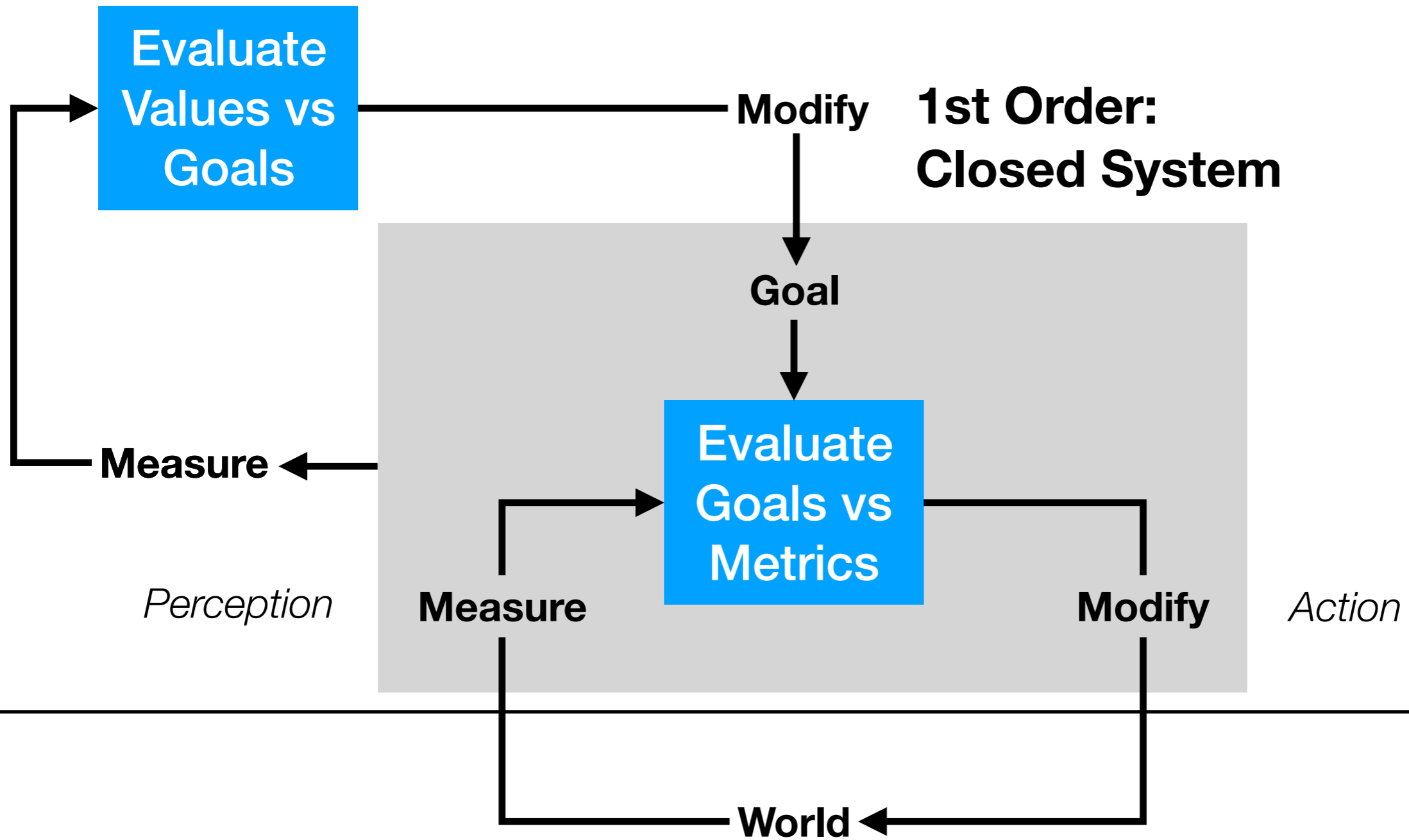


Reinforcement Learning

AI Definition of Intelligence

“Intelligence measures an agent’s ability to achieve goals in a wide range of environments.”

2nd Order: Learning System



Closed System

