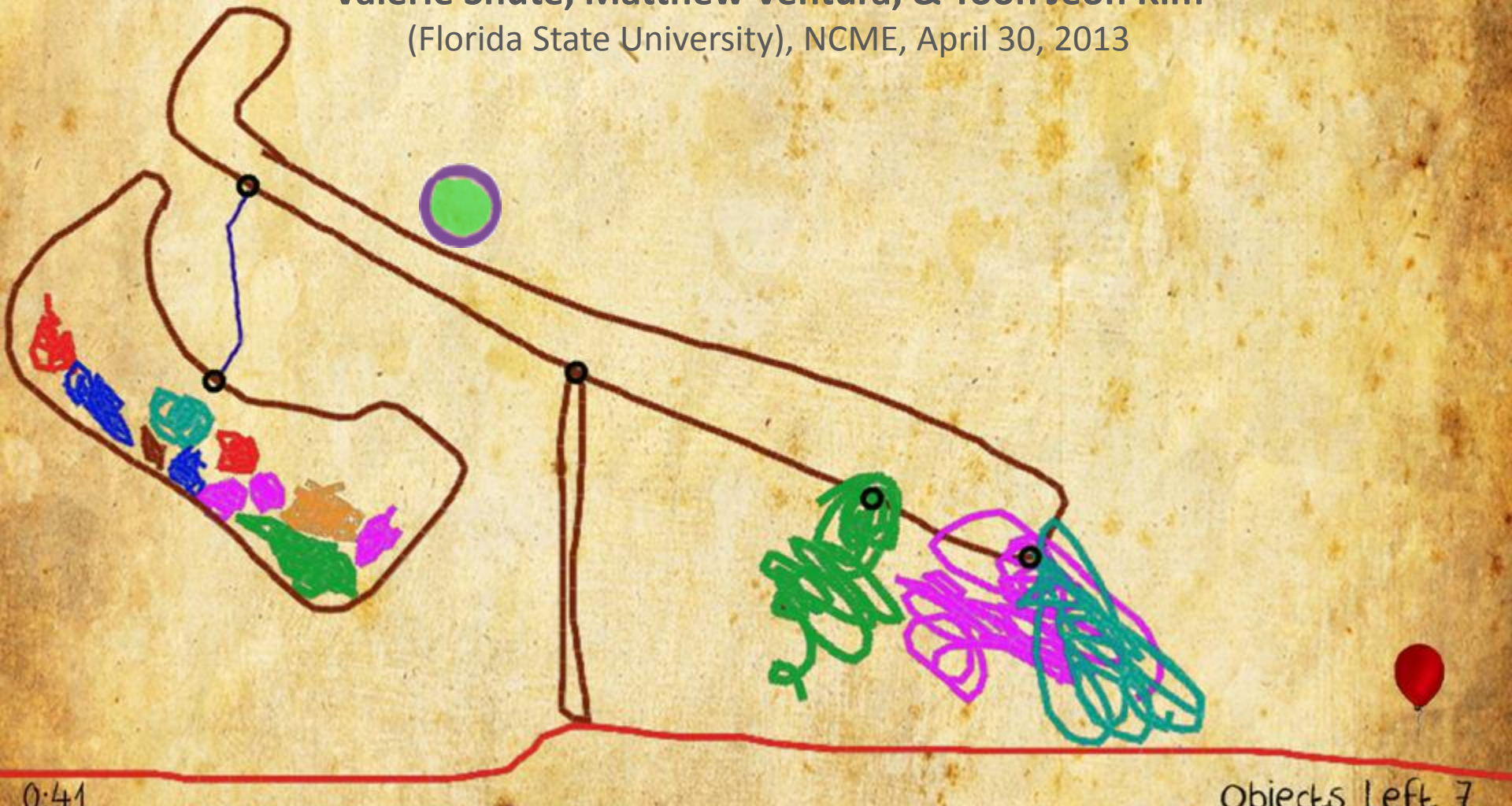
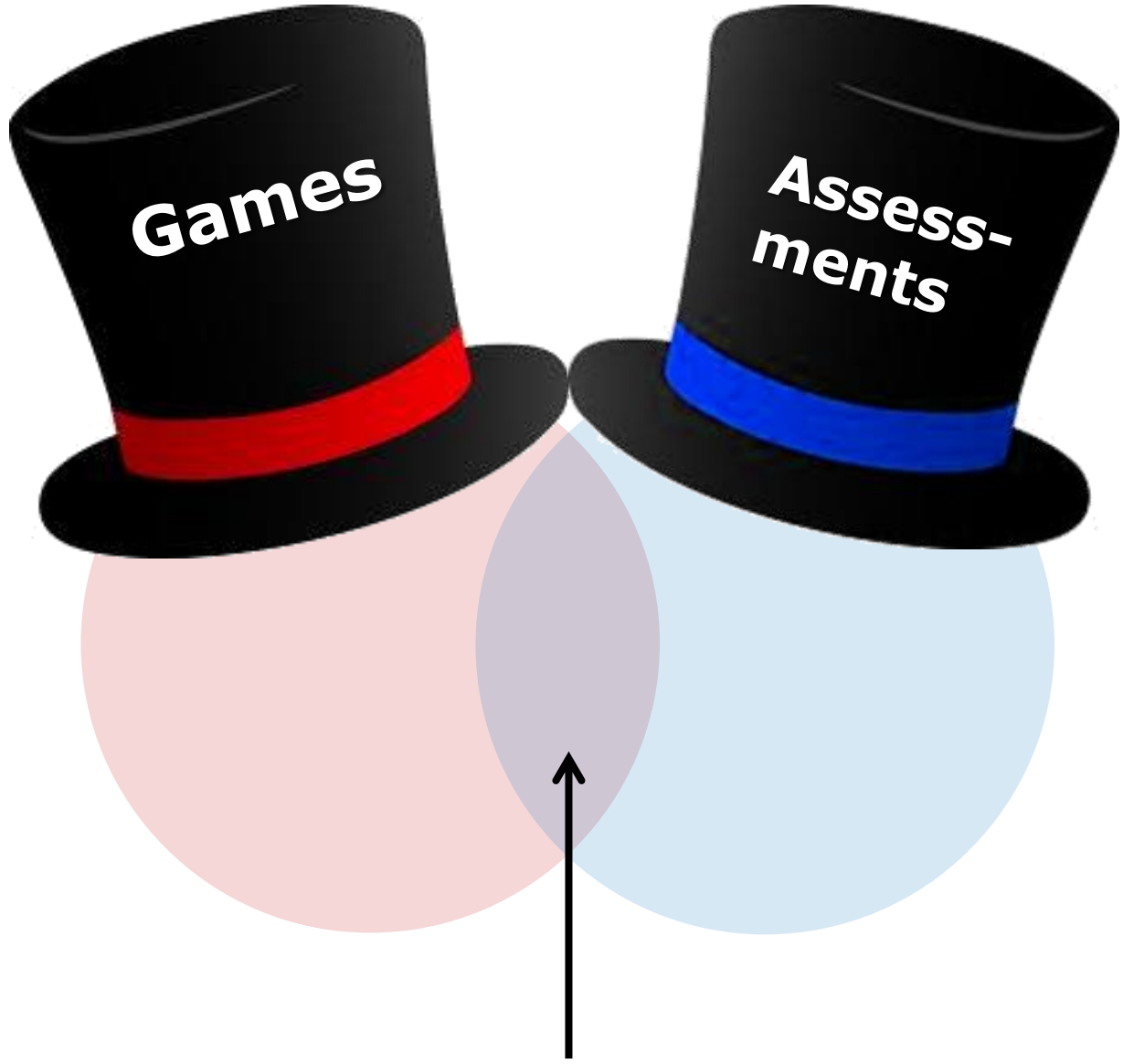
Game-based formative assessment:

Newton's Playground

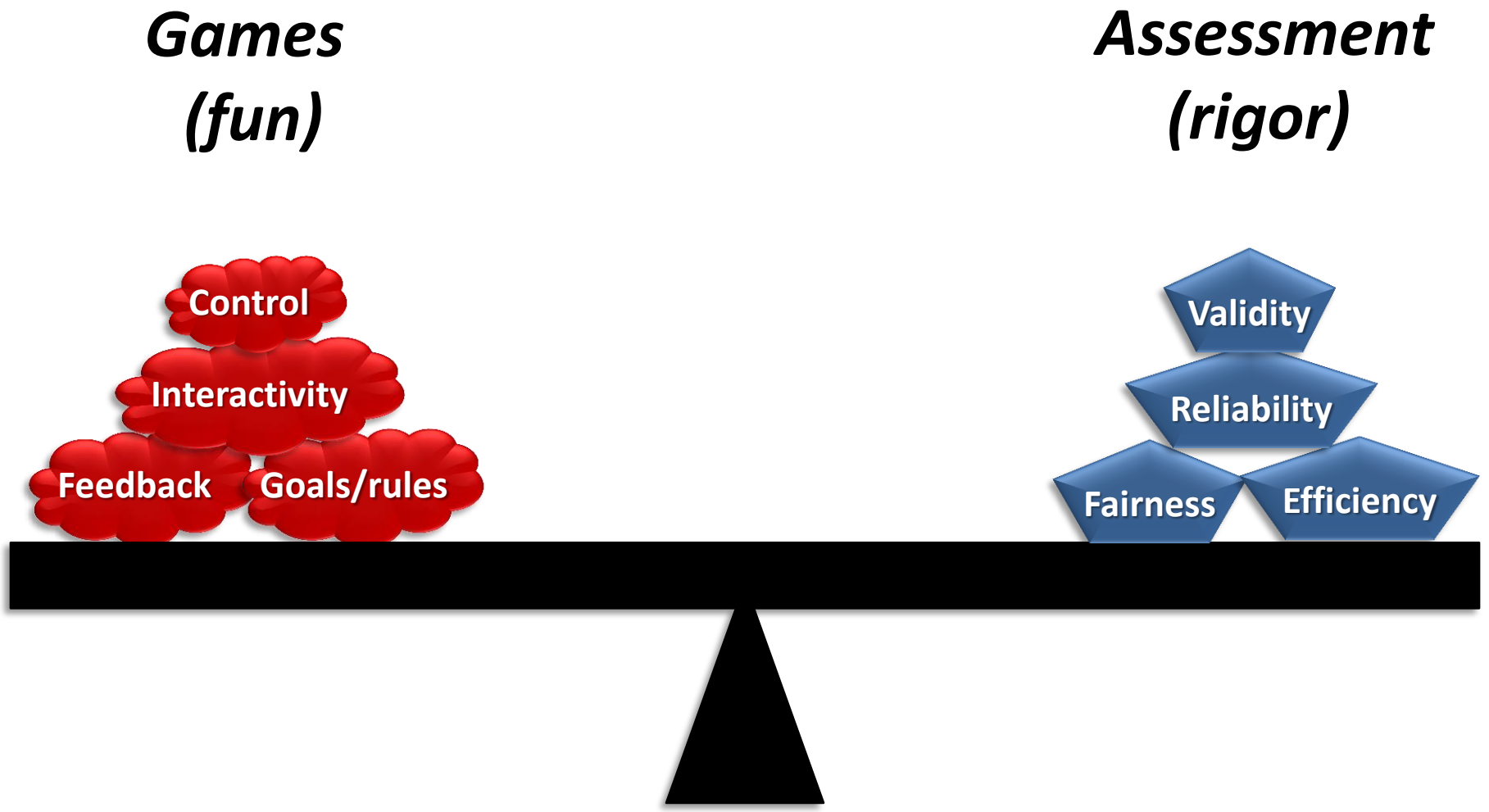
Valerie Shute, Matthew Ventura, & Yoon Jeon Kim
(Florida State University), NCME, April 30, 2013



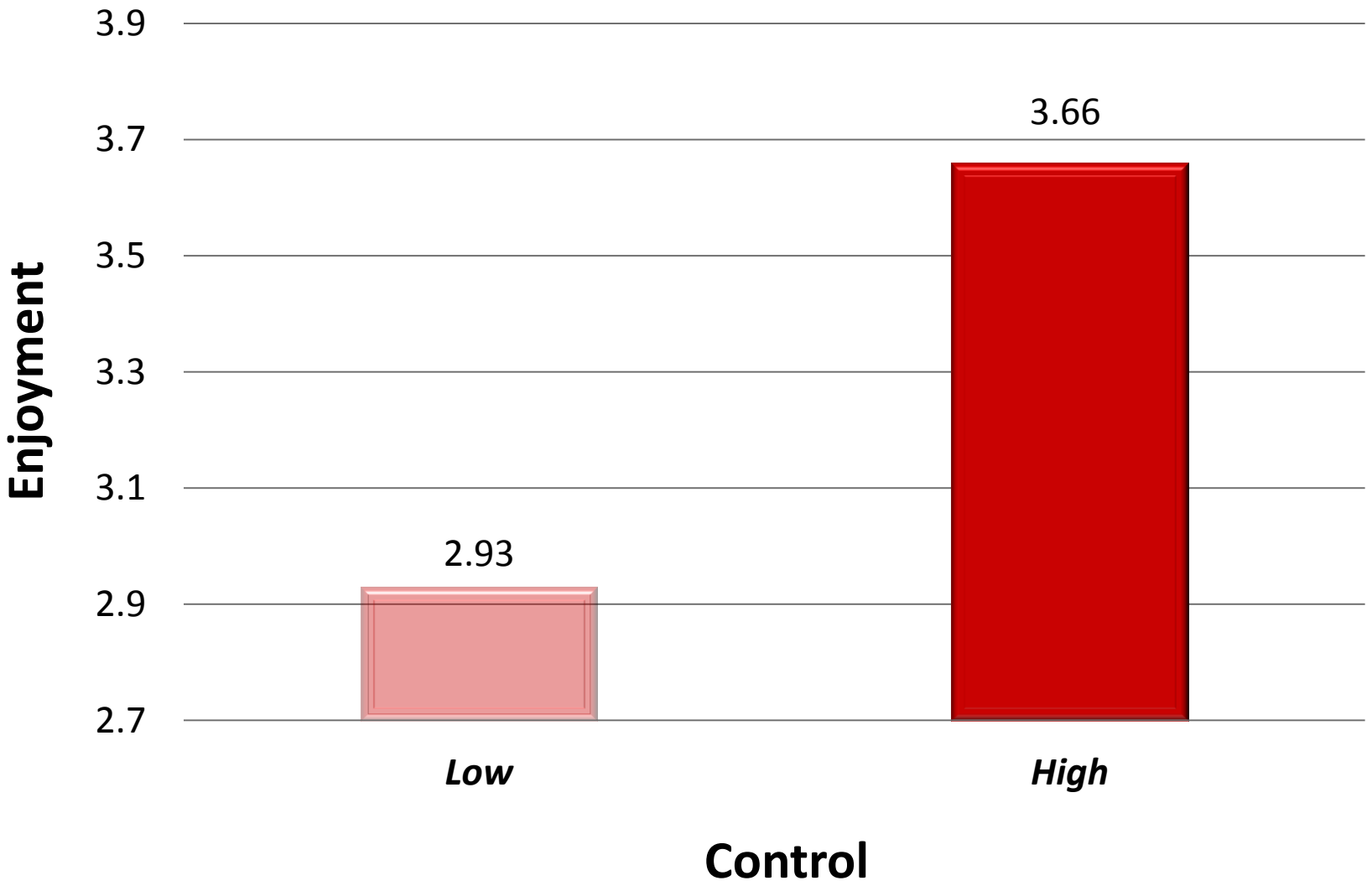

Fun & Games
Assessment Needs



Game-based stealth assessment

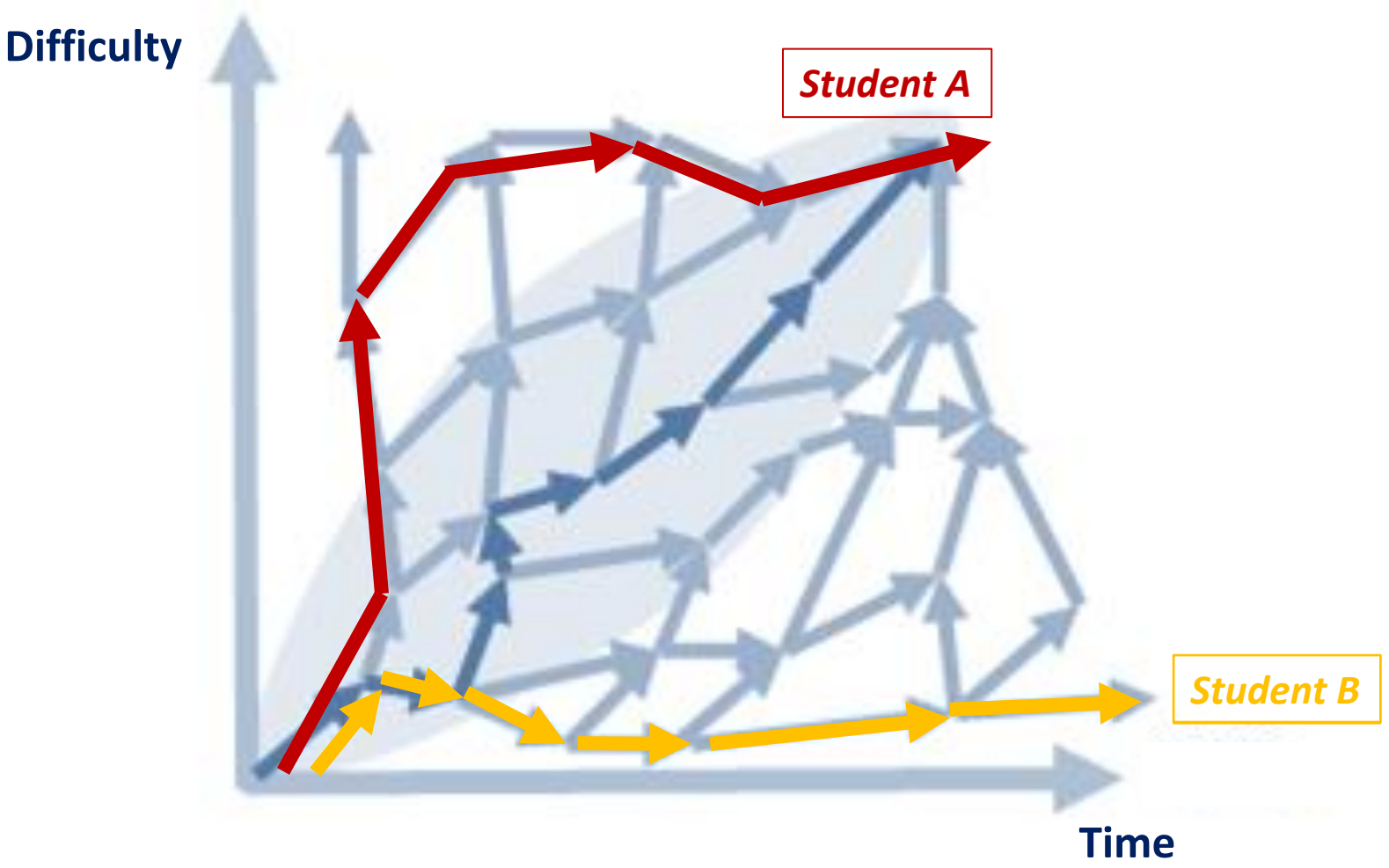

Games
(fun)
Assessment
(rigor)
Control
Interactivity
Feedback
Goals/rules
Validity
Reliability
Fairness
Efficiency

Control & Games



Klimmt, C., Hartmann, T., & Frey, A. (2007). Effectance and Control as Determinants of Video Game Enjoyment. *CyberPsychology & Behavior, 10(*6), 845-848. doi: 10.1089/cpb.2007.9942

Control & Assessment



Feedback & Assessment

QUESTION: When students are given good feedback on their task solutions, does their learning render the assessment less valid, reliable, or efficient?

ANSWER: **No**

SEE: Shute, V. J., Hansen, E. G., & Almond, R. G. (2008). You can't fatten a hog by weighing it—Or can you? Evaluating an assessment for learning system called ACED. *International Journal of Artificial Intelligence and Education*, *18*(4), 289-316.

Stealth Assessment Features

Seamless & Ubiquitous Assessment

When the cook tastes the soup, that's formative; when the guests taste the soup, that's summative.

Formative & Diagnostic

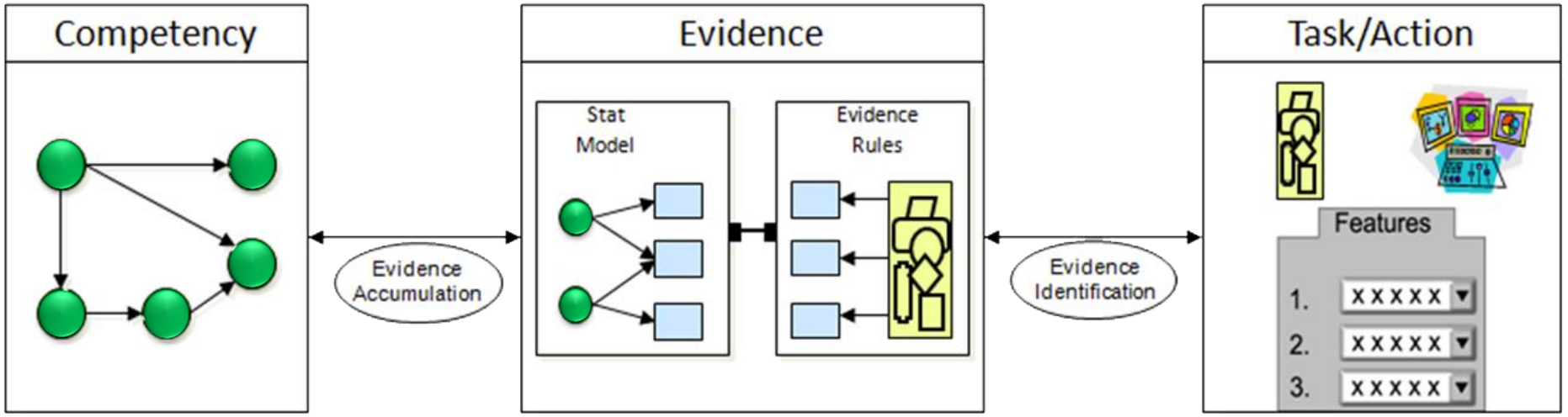
Accurate & Rich Learner Models

Invisible assessment, ***transparent support!***

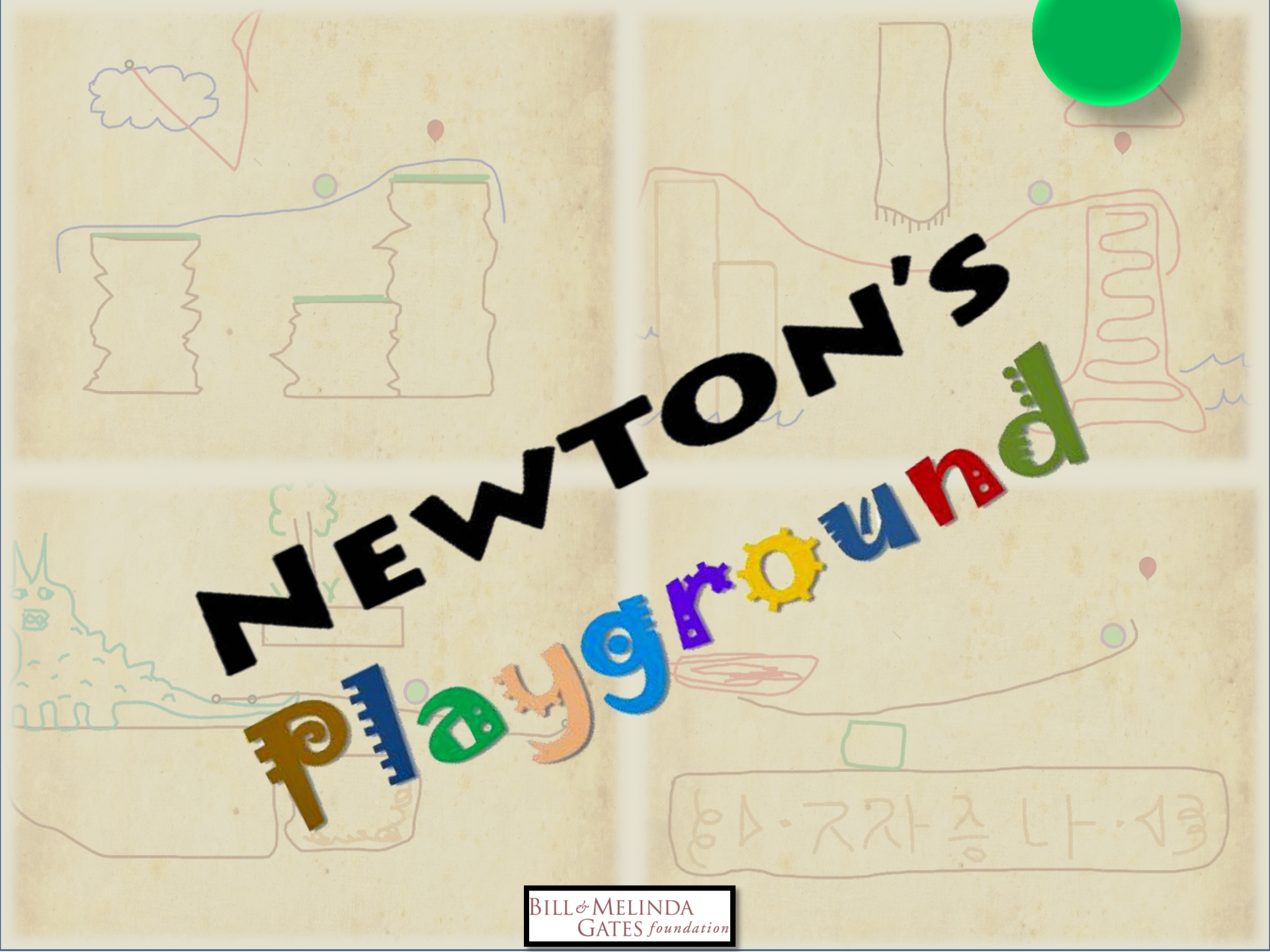
ECD

(e.g., Mislevy, Steinberg, & Almond, 2003)

Assessment Models & Metrics

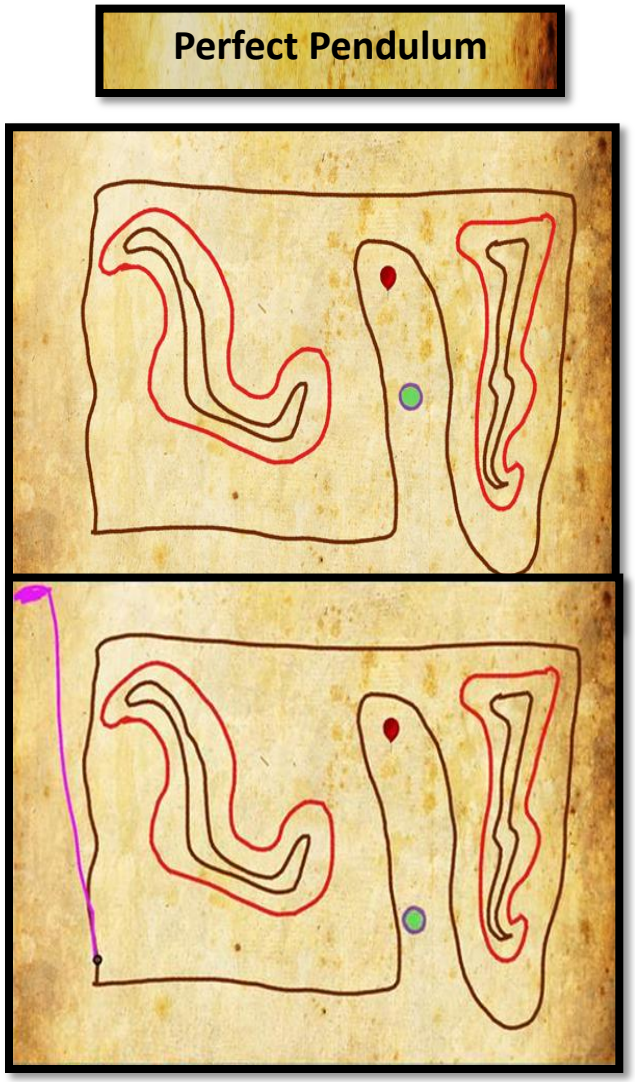


Monitor & Diagnose Success


NEWTON'S
playground
BILL & MELINDA GATES foundation

Newton’s Playground

- ✓ Goal: guide a  to a  . Everything obeys basic rules of physics (e.g., gravity, Newton's three laws of motion).
- ✓ Player draws physical objects that "come to life" when drawn (e.g., levers, ramps, pendulums) to get ball to balloon.
- ✓ Players can solve problems in many different ways, striving for the ***awesomest*** one.

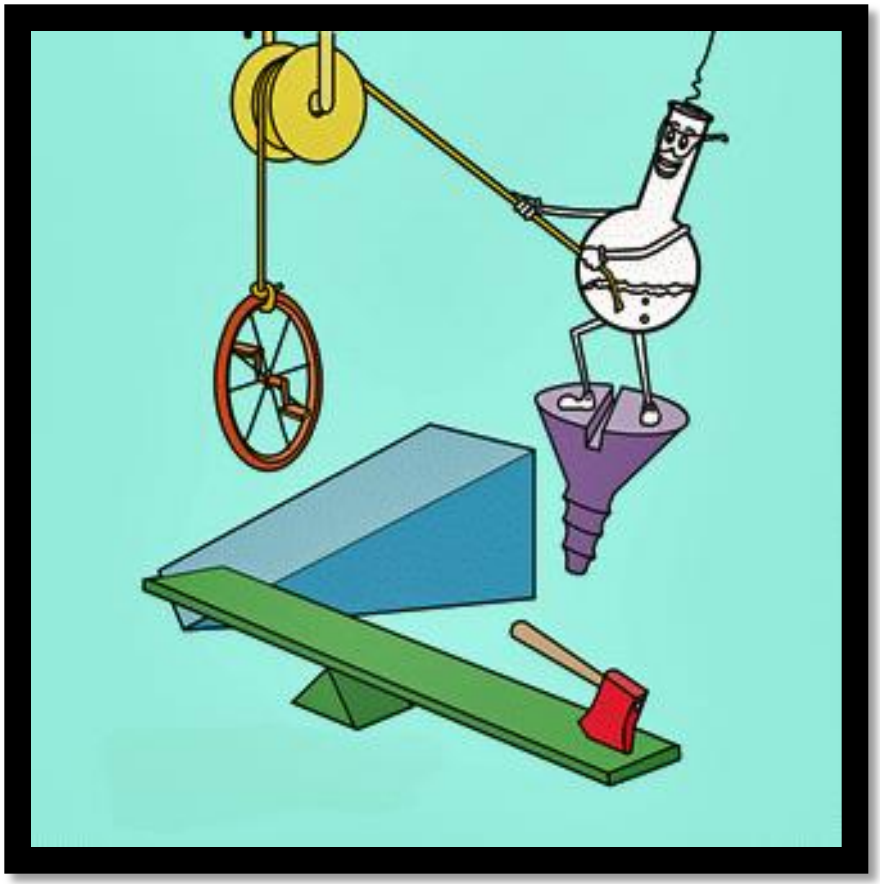


Qualitative Physics

(Ploetzner, VanLehn, 1997)

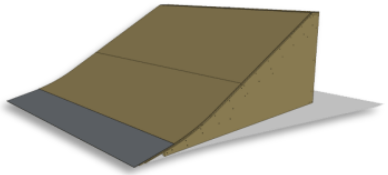
Nonverbal understanding of:

1. Newton's three laws of motion
2. Balance
3. Mass
4. Gravity

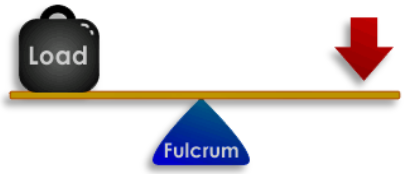


Agents of Force/Motion

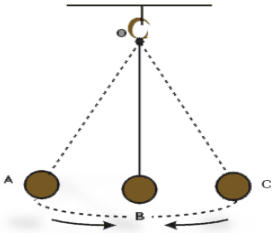
- ***Ramp***: Used to change the direction of the motion of the ball (or another object).



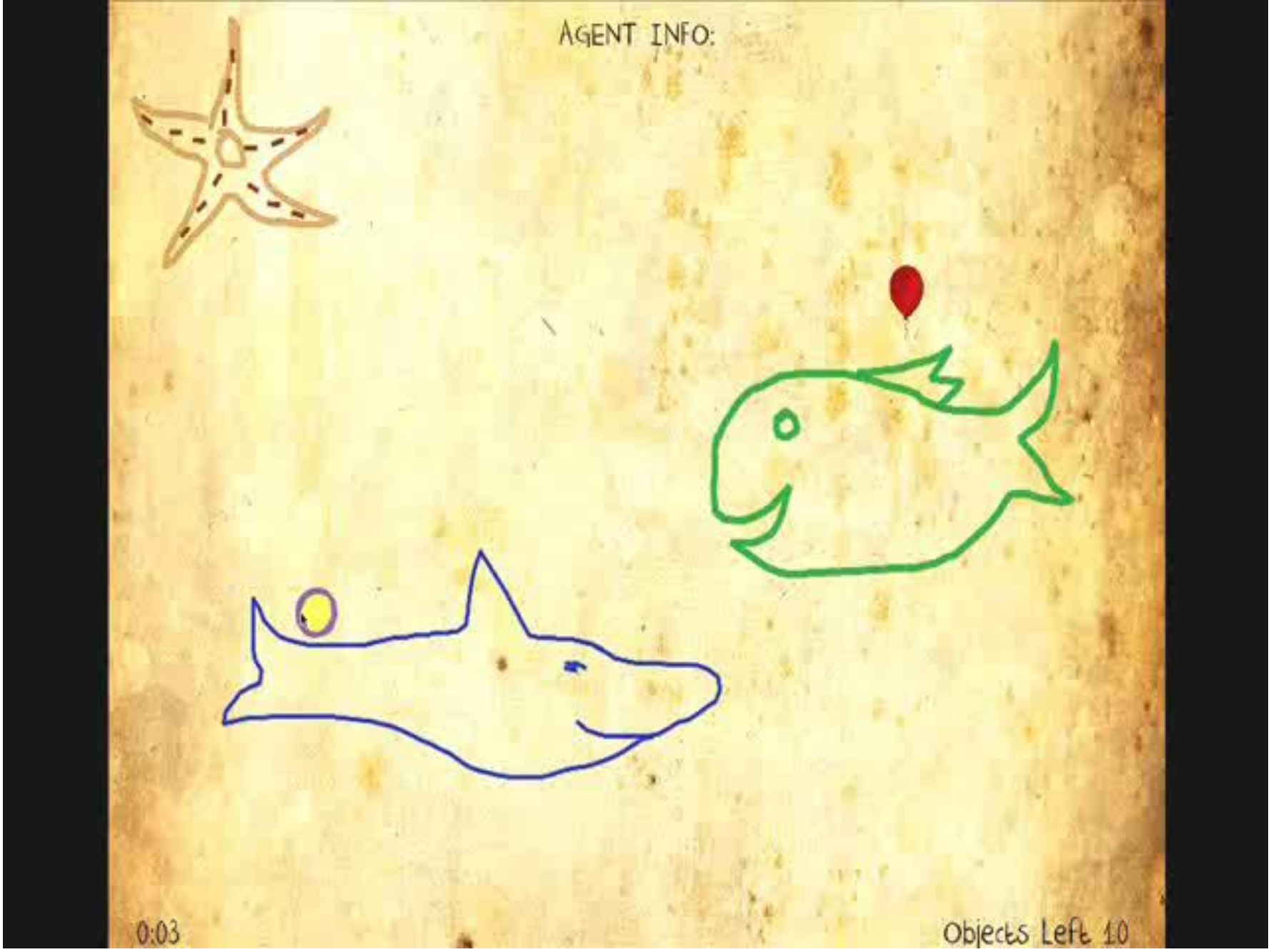
- ***Lever***: Rotates around a fixed point usually called a fulcrum or pivot point.



- ***Pendulum***: Directs an impulse tangent to its direction of motion. Secured at the top by a pin.



- ***Springboard***: Stores elastic potential energy from falling weight; becomes kinetic as weight is released.

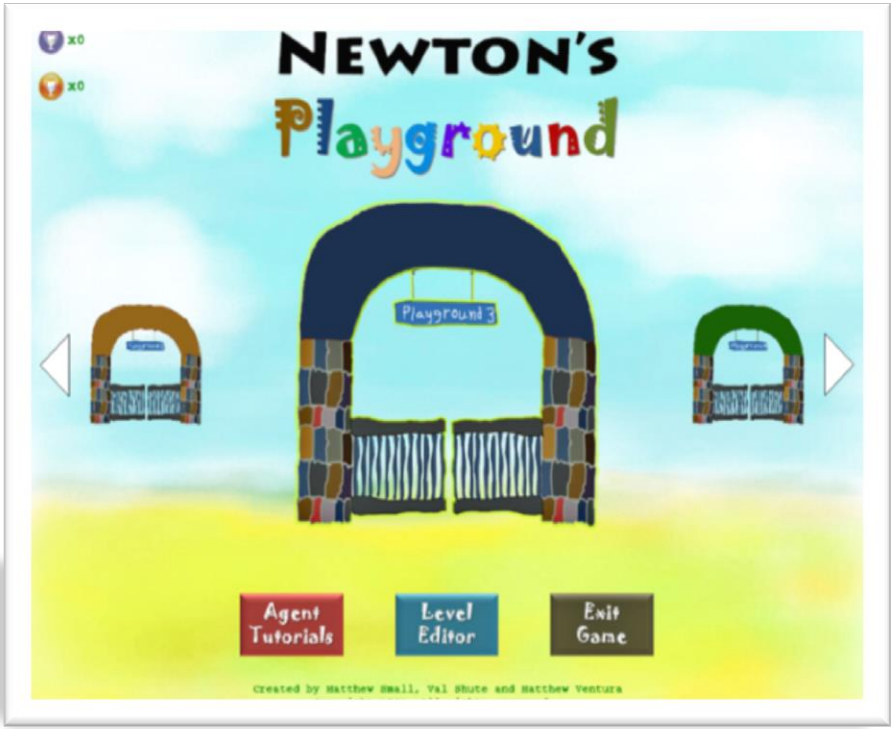

AGENT INFO:
0:03
Objects Left 10

Difficulty Indices

- ***Relative location of ball to balloon***. If balloon is above ball, forces player to use lever, springboard, or pendulum to solve the problem (0-1).
- ***Obstacles***. If pathway between ball and balloon is obstructed, player must project ball in specific trajectory (0-2).
- ***Distinct agents of force/motion***. A problem may require one or more agents to get ball to the balloon (0-1).
- ***Novelty***. A problem is not like any other problems played so solution is not easily determined from prior experiences (0-2).

Game design choices in NP

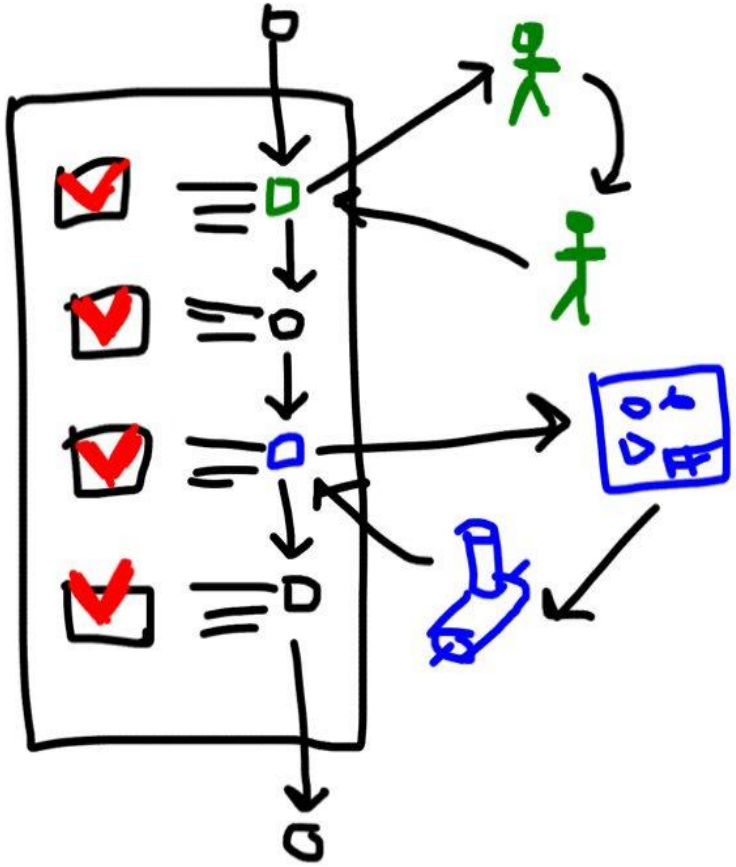
- ***Control***: Freedom to play any problem anytime (set up in playgrounds of increasing difficulty)
- ***Interactivity***: Create their own responses; multiple valid solutions



- ***Feedback***: Gold vs. silver trophies.
- ***Goals/rules***: super clear (get ball to balloon)

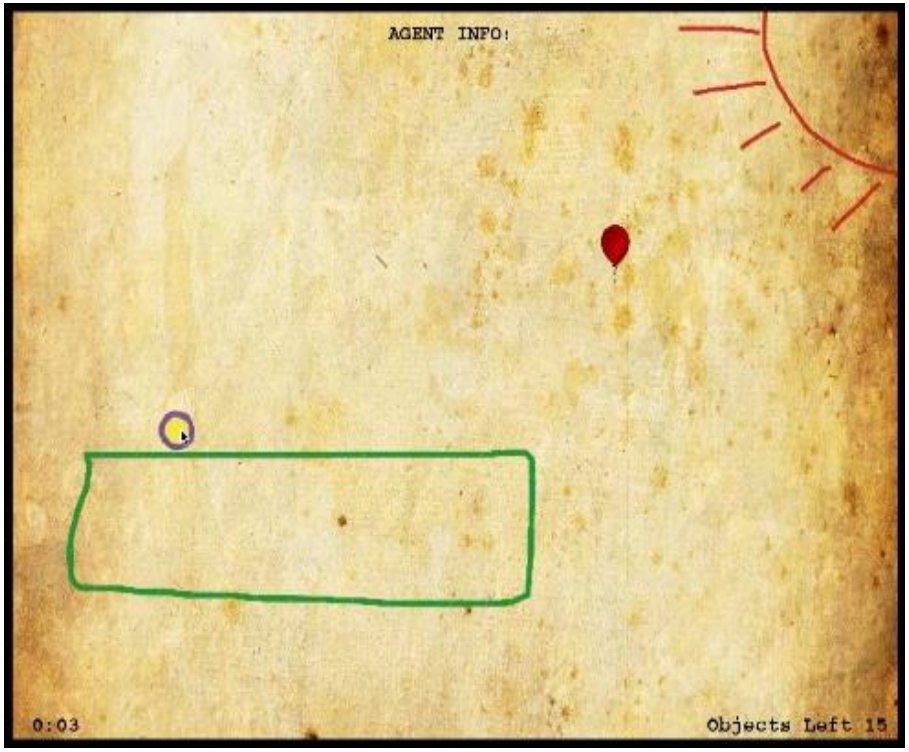
Task-level design choices

- Balance evidence elicitation
  - » All agents used
  - » Playgrounds balanced
- Focus evidence
  - » Some levels target just 1 agent (e.g., pendulum only)
- Increase difficulty (Playgrounds 1-7)
  - » Discrimination
- Don't suck out the fun
  - » Construction of colorful responses
  - » Variation of challenges

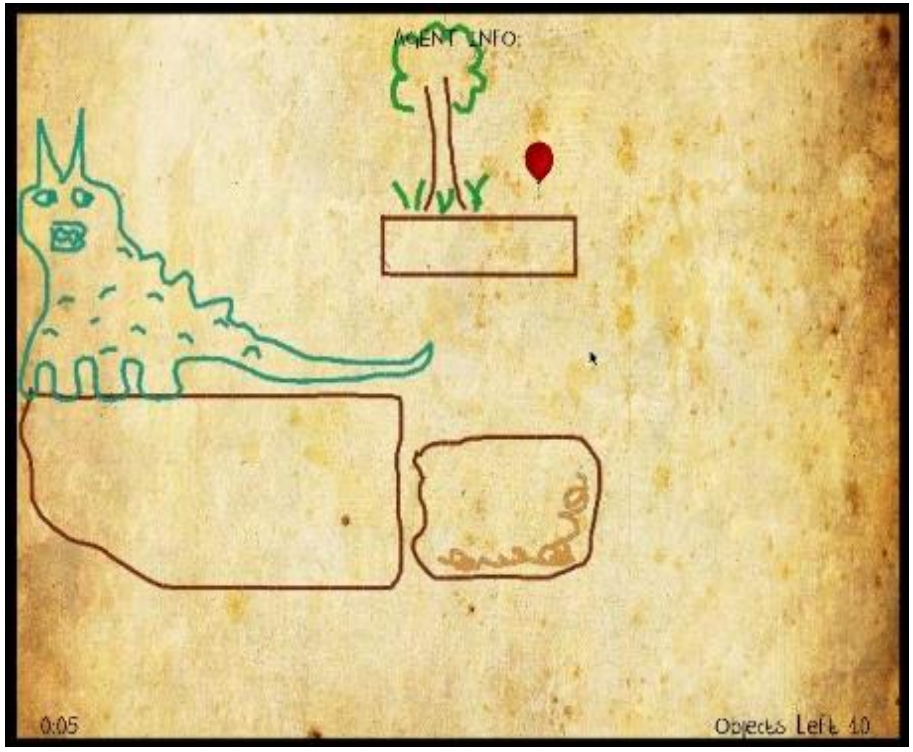


Springboard: Difficulty

Sunny Day: Easy SB

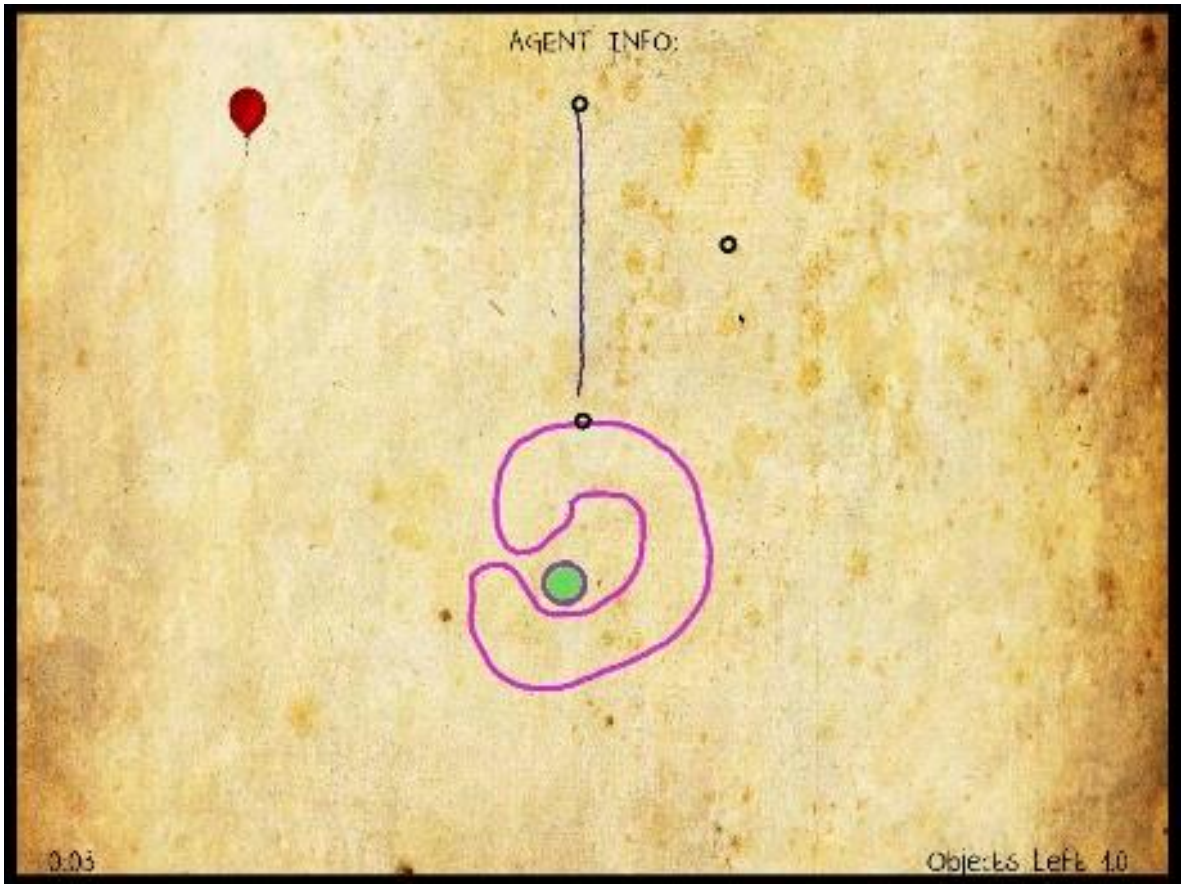


Jurassic Park: Medium SB

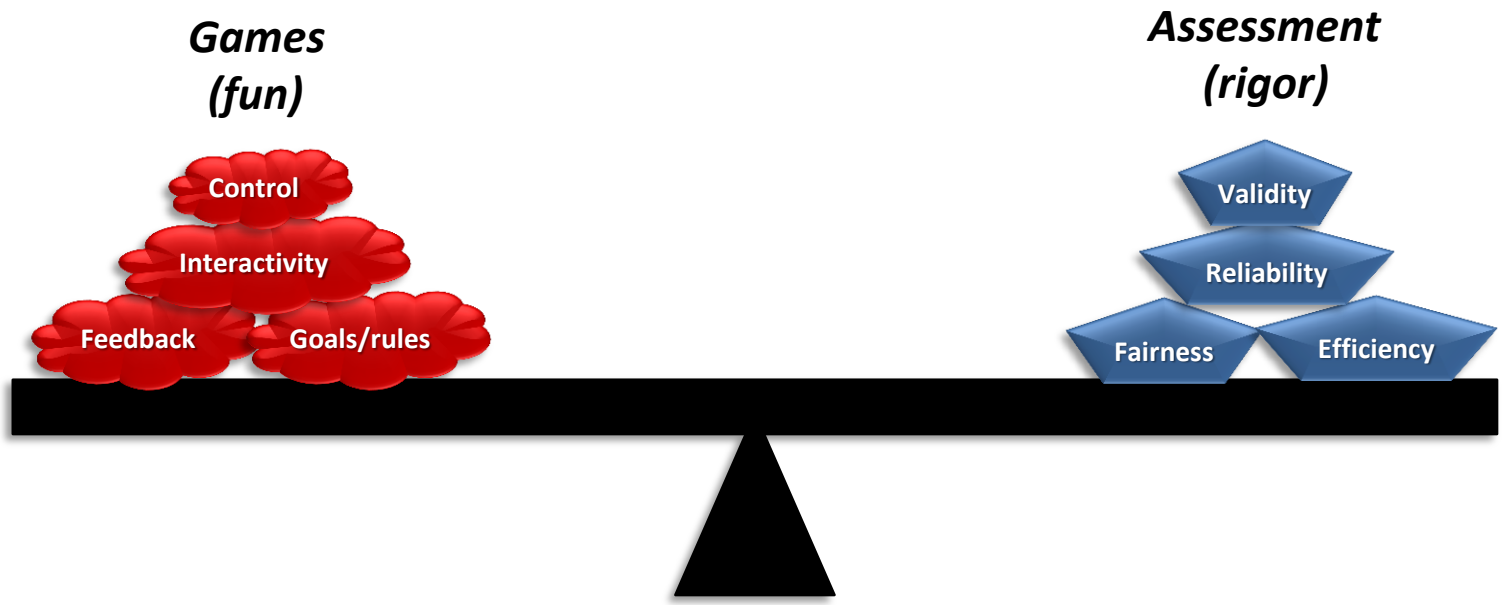


Pendulum problem

Used features of the game task to (subtly) constrain players' choice of agent



How did our game-design decisions affect the quality of the assessment, learning, and enjoyment?



Construct Validity: External & In-game Physics (*N* = 166)

External measure of physics knowledge (pretest) correlated with in-game measures of mastery (number gold trophies per agent).

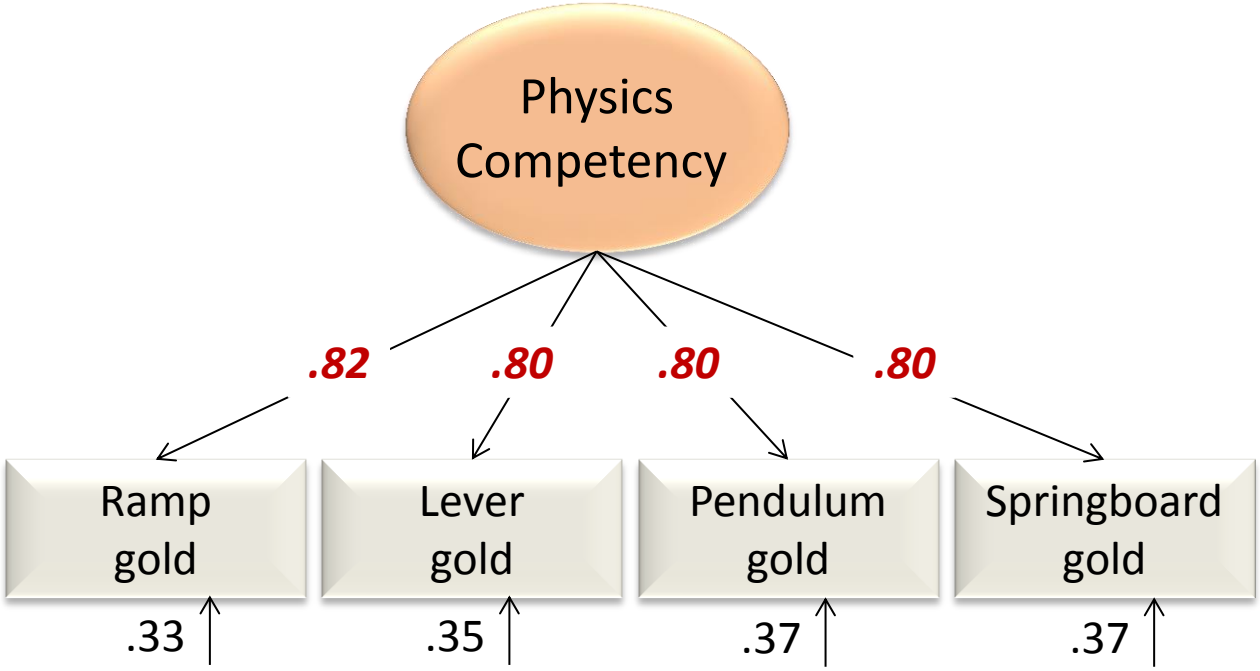
Correlations: Pretest Scores and NP Trophies

Posttest**	0.60
Ramp-silver	0.09
Lever-silver	-0.04
Pendulum-silver	-0.02
Springboard-silver	0.15
Ramp-gold**	0.24
Lever-gold**	0.23
Pendulum-gold**	0.34
Springboard-gold**	0.41

N = 166; ** $p < .01$

Results: Construct Consistency

1. CFA – Gold trophies by four agents: $X^2/df < 3$, CFI > .95, RMSEA < .05, SRMR < .05

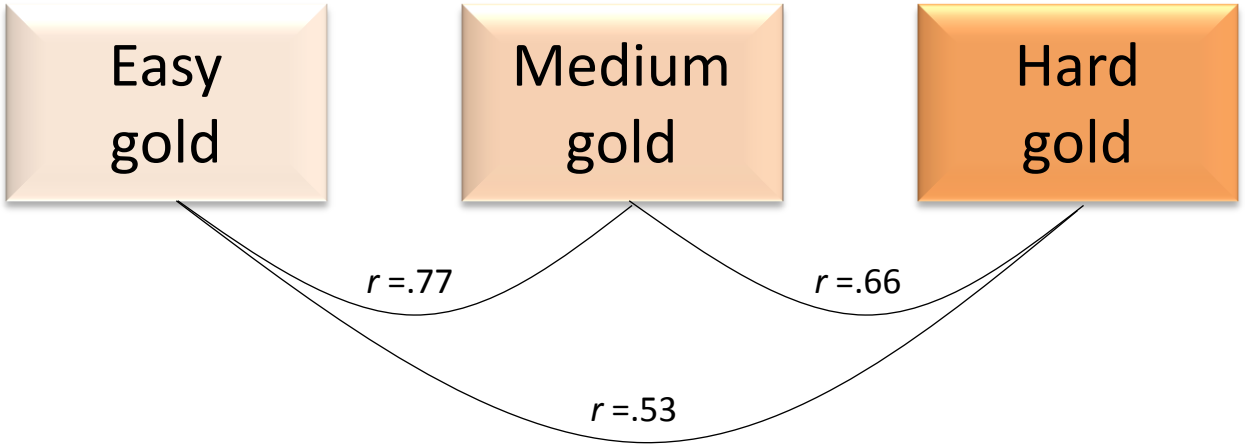


2. Intraclass correlation = ***.85*** (Ramp, Level, Pendulum, Springboard gold trophies)

3. Pairwise correlations: RxL = .67; RxP = .64; RxS=.66; LxP=.64; LxS=.63; PxS=.65

Results: Construct Consistency

1. Intraclass correlation = ***.82*** (Easy, Medium, Hard gold trophies)



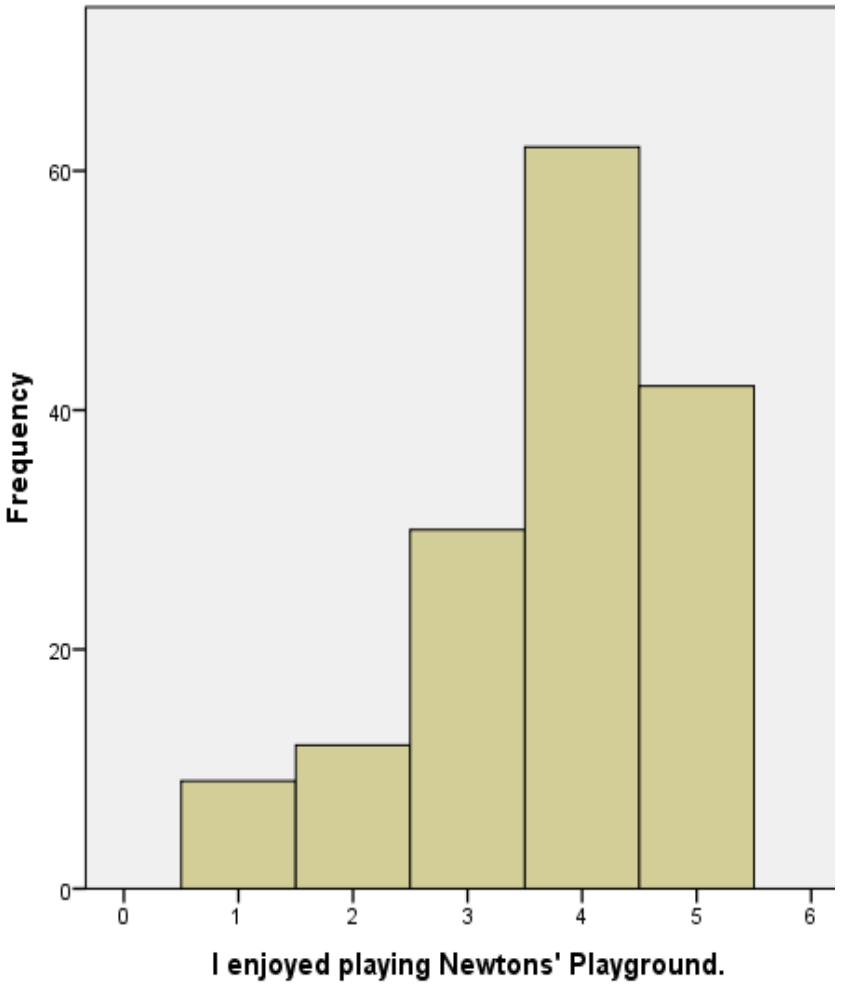
2. Cronbach's alpha = ***.87***

Data: gold trophy info (NA, 0, 1)
Valid Cases: 110 (out of 169)
Levels: 29 (out of 74)

Results: Learning & Fun

How did the decisions work out?

- ***Learning***: Significant difference between pretest & posttest scores: $F\ (1, 153) = 4.24;\ p < .05$ simply after 4 hr gameplay.
- ***Enjoyment***: Kids enjoyed the game (1=dislike; 5=like; *M*=4, *SD* = 1). Males & females enjoyed equally (after controlling for pretest).



Next Steps: Formative Assessment

- Info on competencies used by (a) ***teachers*** (to adjust instruction & give good feedback), (b) ***students*** (to reflect on how they're doing), and (c) ***system*** (to select new gaming experiences), such as:
  - Present problem requiring agents not mastered
  - Provide hints re: agent solutions
  - Give rewards for novel agent use
  - Include formalizations (and values) in simulation (e.g., level editor)
  - Display current estimates of competency levels in NP (progress indicators) so students act to improve them.
- Develop curriculum to wrap around game—lesson plans, activities (e.g., student levels demo'ed and discussed in class), etc.

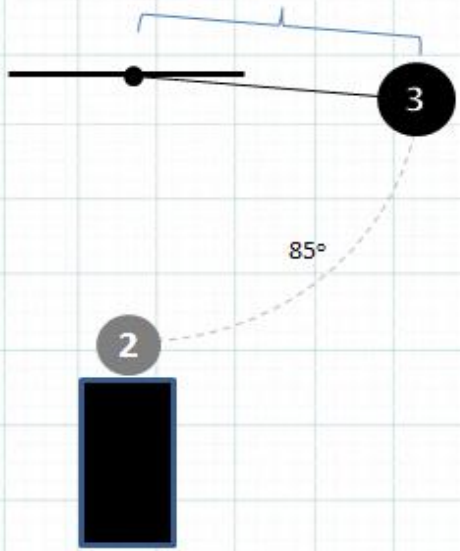
Thank you!

Questions?

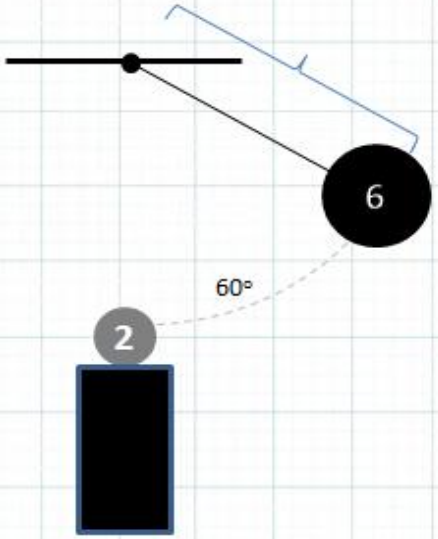
Email: vshute@fsu.edu
Website: http://www.myweb.fsu.edu/vshute
Download NP: http://www.gameassesslearn.org/newton/

Physics Test

A

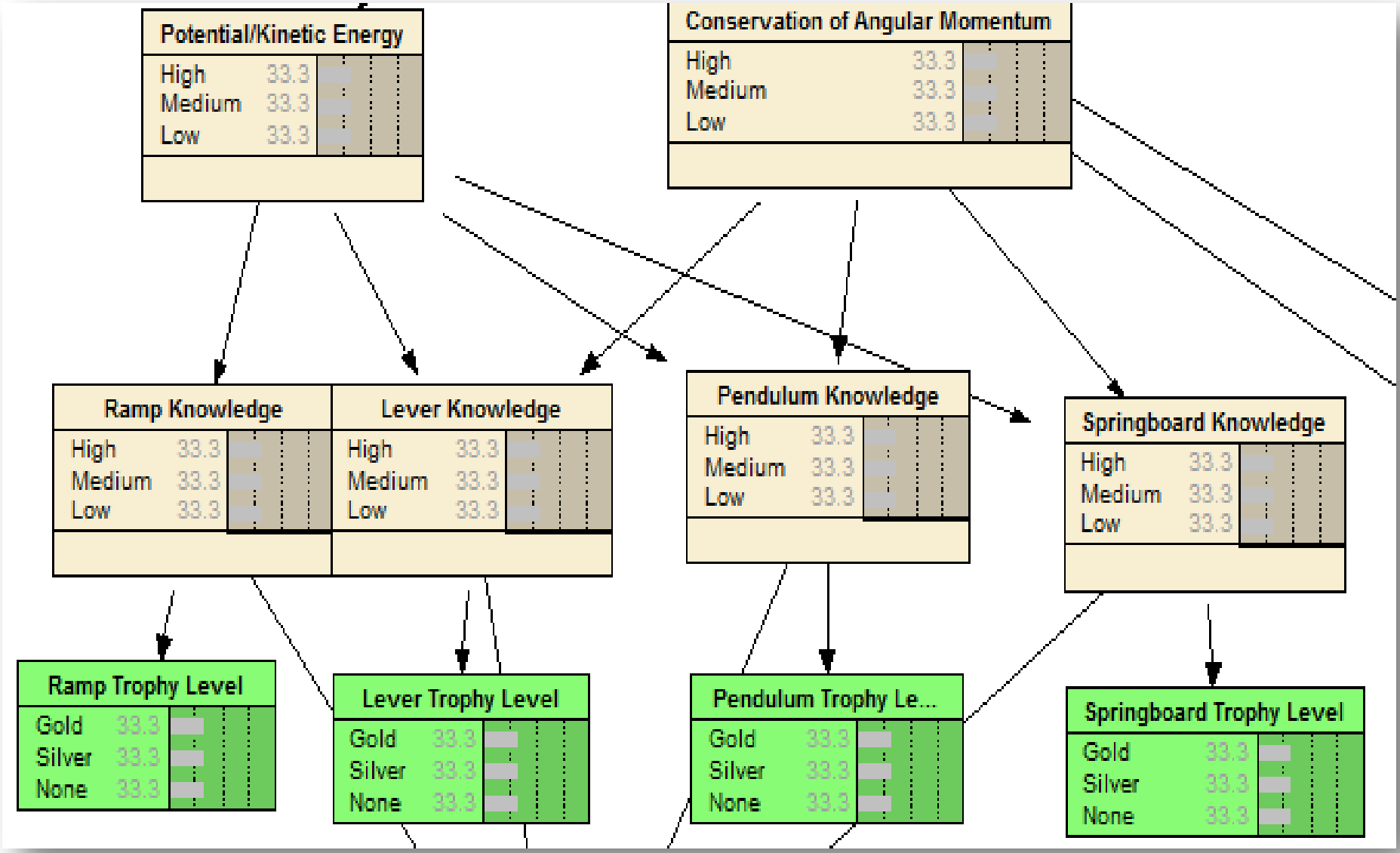


B

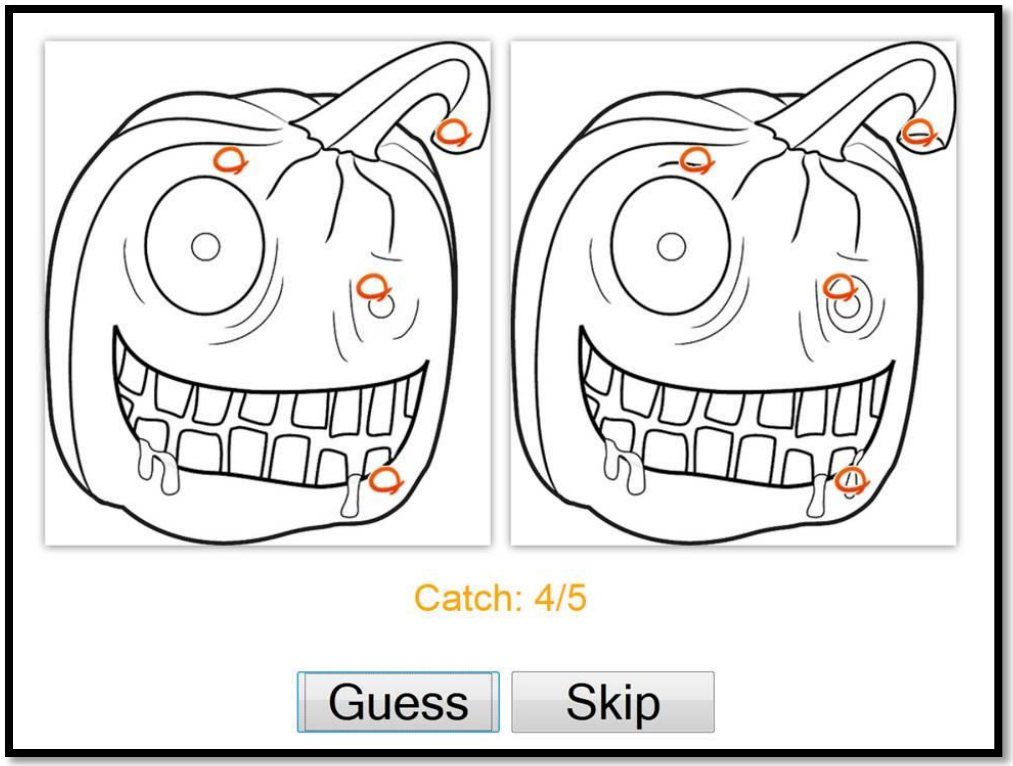


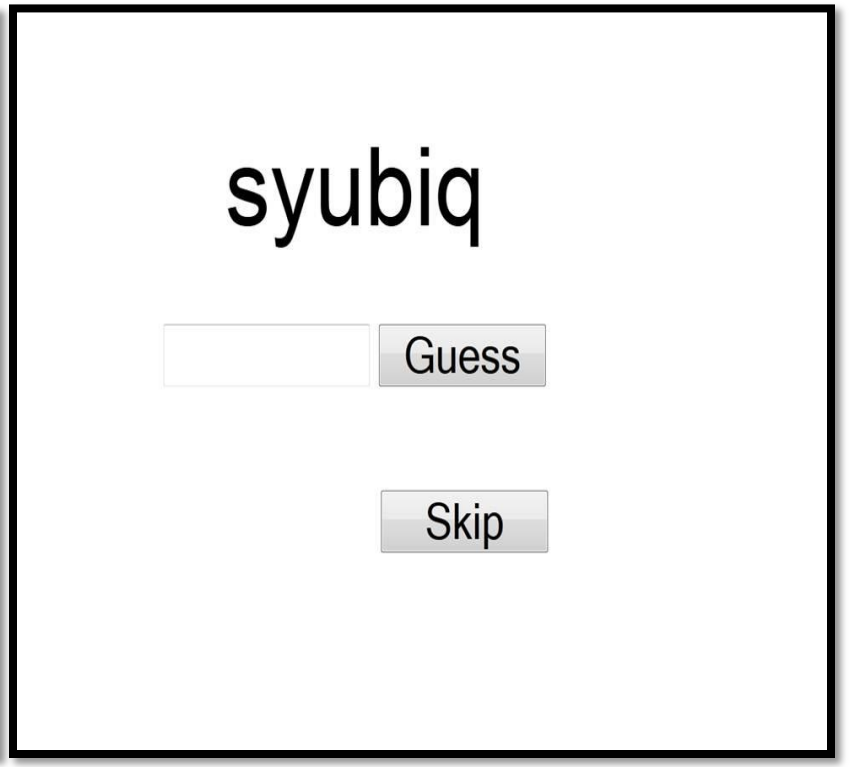
The pendulums are swinging down to hit gray balls. The pendulums each have the same length, but they start their swings from different angles and they have different masses. In which figure will the gray ball travel the fastest after being hit by the pendulum?

a) A
b) B
c) The gray balls in A and B will move at the same speed after being hit
d) I do not know


Potential/Kinetic Energy
High 33.3
Medium 33.3
Low 33.3
Conservation of Angular Momentum
High 33.3
Medium 33.3
Low 33.3
Ramp Knowledge
High 33.3
Medium 33.3
Low 33.3
Lever Knowledge
High 33.3
Medium 33.3
Low 33.3
Pendulum Knowledge
High 33.3
Medium 33.3
Low 33.3
Springboard Knowledge
High 33.3
Medium 33.3
Low 33.3
Ramp Trophy Level
Gold 33.3
Silver 33.3
None 33.3
Lever Trophy Level
Gold 33.3
Silver 33.3
None 33.3
Pendulum Trophy Le...
Gold 33.3
Silver 33.3
None 33.3
Springboard Trophy Level
Gold 33.3
Silver 33.3
None 33.3

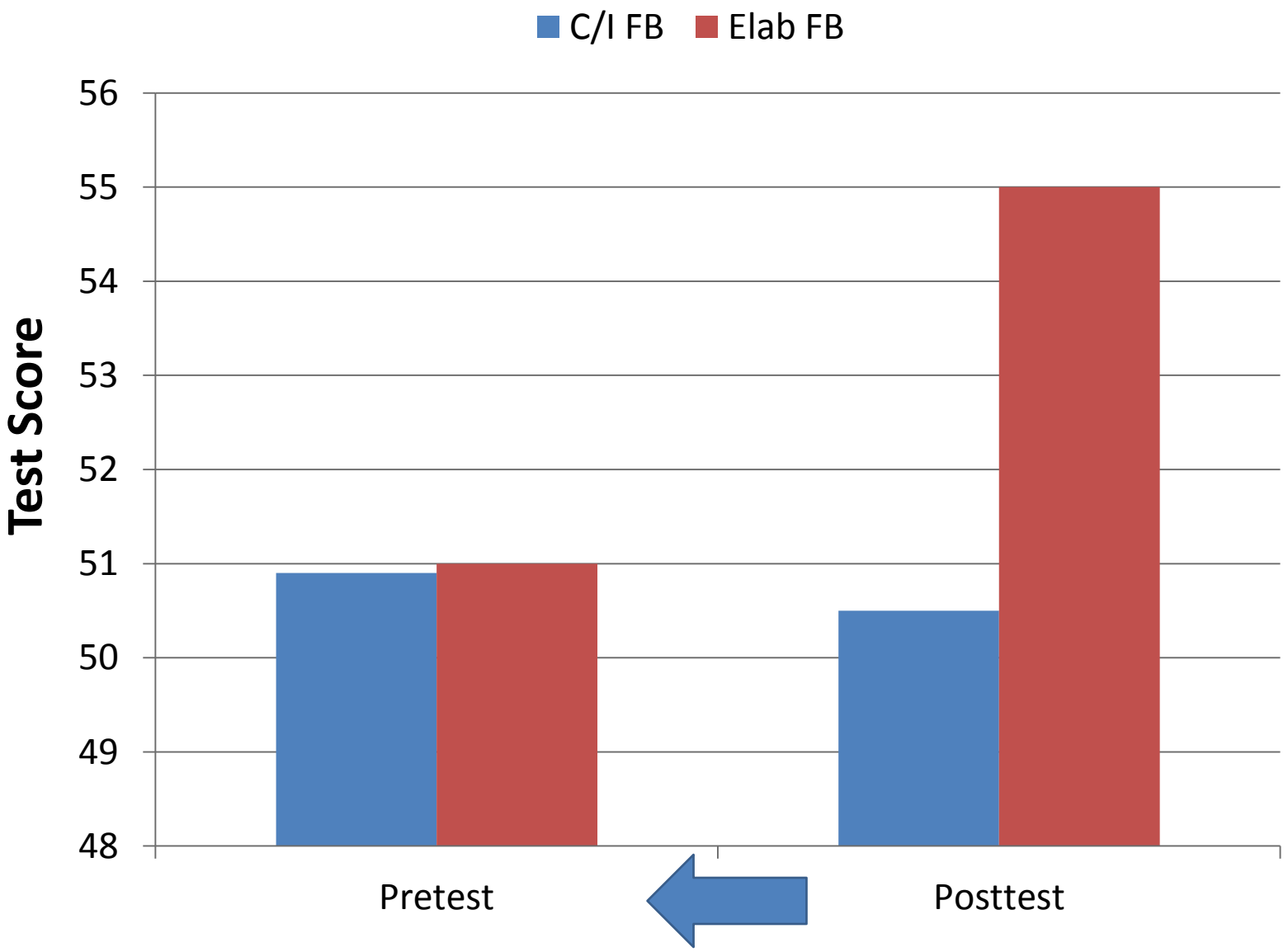
Persistence Test





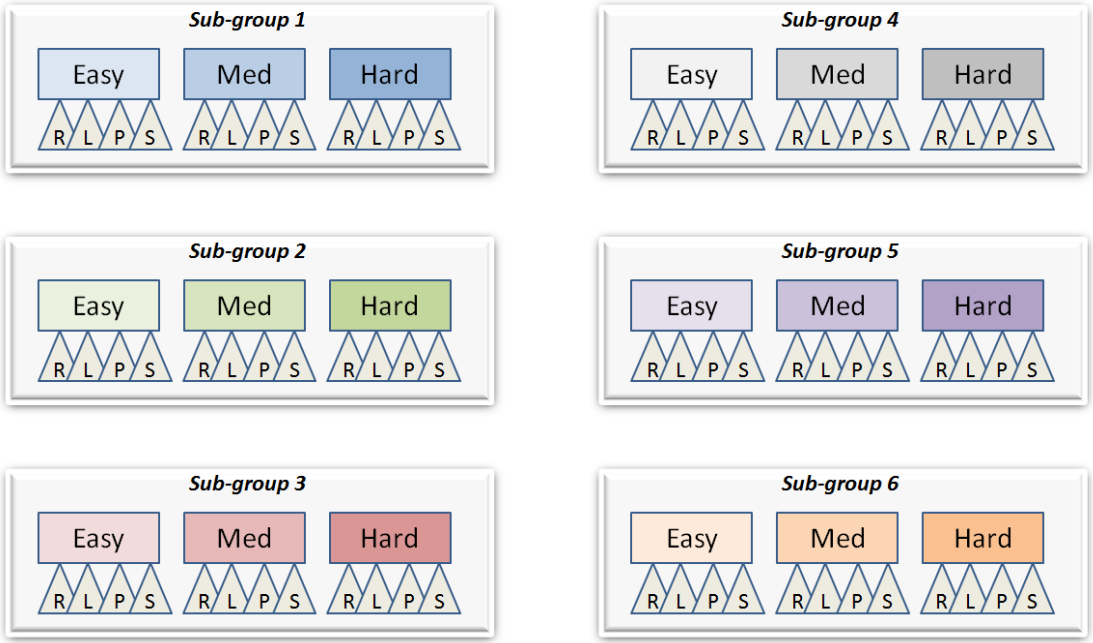
VALIDATION OF THE MEASURE Ventura, M., Shute, V. J., & Zhao, W. (2012). The relationship between video game use and a performance-based measure of persistence. *Computers & Education, 60,* 52-58.

Feedback in AfL System



Jackknife Variance Estimation (Consistency of assessment)

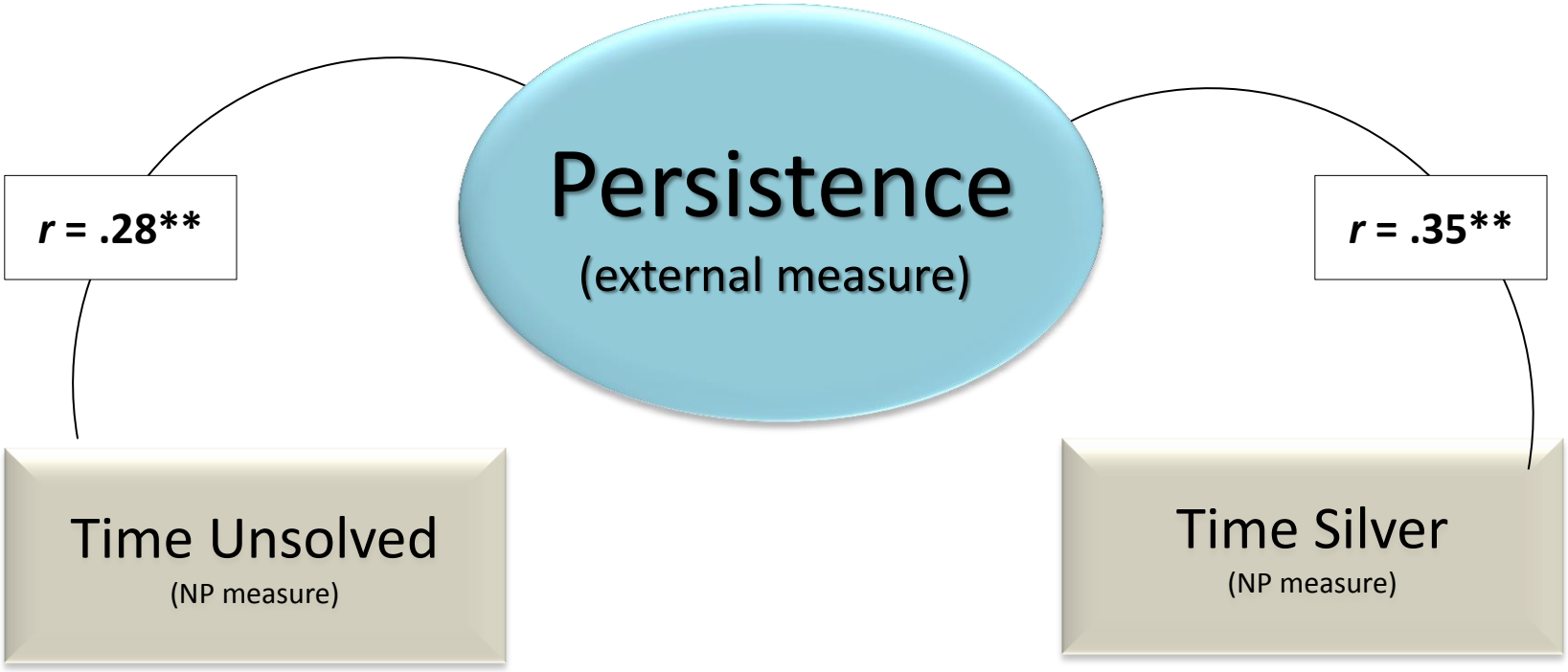
- Jackknife resampling: Compared variance of full sample (74 levels) with variance caused by different task formats (i.e., levels)



- Used gold trophy information (NA, 0, and 1)

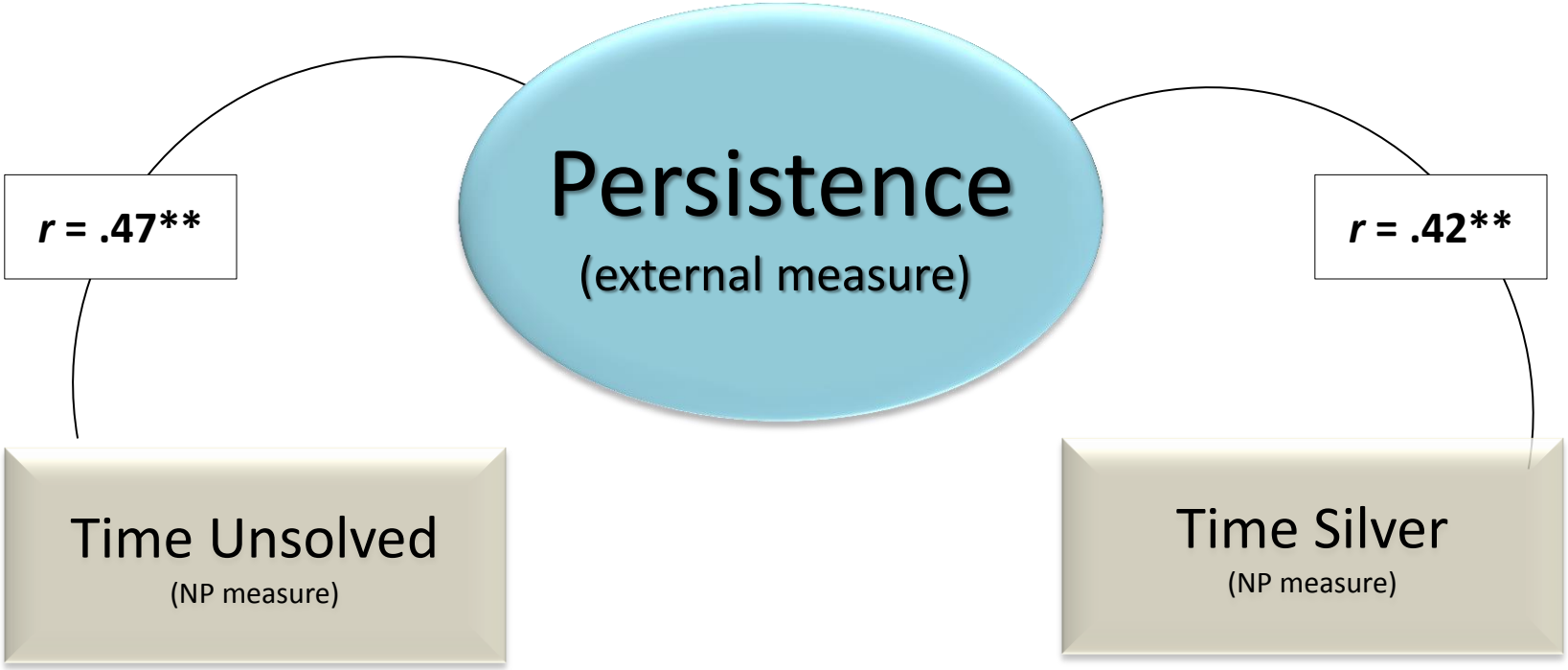
- JK variance (1.1) divided by full sample variance (77.57) = 0.015; reliability = .985!

Convergent Validity: Persistence



Time on Gold Trophies (*r*=.07)

Convergent Validity: Persistence (just low performers)



Time on Gold Trophies (*r*= .004)

Can there be validity without reliability?

(Moss, 1994)

"Although the focus here is on reliability (consistency among independent measures intended as interchangeable), it should be clear that reliability is an aspect of construct validity (consonance among multiple lines of evidence supporting the intended interpretation over alternative interpretations). And as assessment becomes less standardized, distinctions between reliability and validity blur. "