Analytics

What's inside LON-CAPA data?

Gerd Kortemeyer

Michigan State University

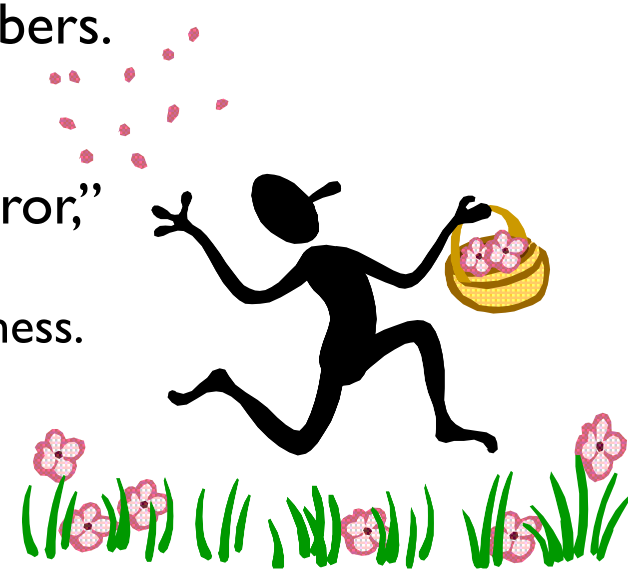
LON-CAPA Conference
2017

… or

- all the questions you wish you had not asked

Apologies

- Presenting a garden variety of results from several years
- Not a coherent story, instead just giving an overview of what might be in the data
- There are many, many null-results, but not presented here
  - There is a lot of noise in the data that does not go away with bigger numbers.
    - Big Data does not suddenly make everything clear-cut
  - This noise is not measurement "error," it's weird behavior of students.
    - Big Data sometimes just more weirdness.
    - Students are not particles.



Analyzing online course components

- We are analyzing online course components in a variety of scenarios:
  - MOOCs
  - Virtual University Courses
  - Blended Courses
  - Flipped Courses
  - Online Textbooks
  - …

Quite a lot of data, actually …

Data in LON-CAPA

- 160 partner institutions
- 48% postsecondary institutions
- 440,000 shared learning objects
- 198,000 shared homework problems
- 7,700 courses hosted since 1999
- 965,000 student-course enrollments served since 1999
- 94% postsecondary student-course enrollments
- 150,000 student-course enrollments per year
- 73,520,000 problems served since 1999
- 138,320,000 problem transactions since 1999
- 72,560,000 problems solved since 1999

Unproductive Behaviors

- Unproductive behaviors
  - Selective reading – only studying a subset of the materials
  - Cramming – studying "last minute"
  - Guessing – entering random solutions, not thinking
  - Copying – copying solutions from other students
- Cannot be observed with traditional textbooks and courses, but can be measured in online course components

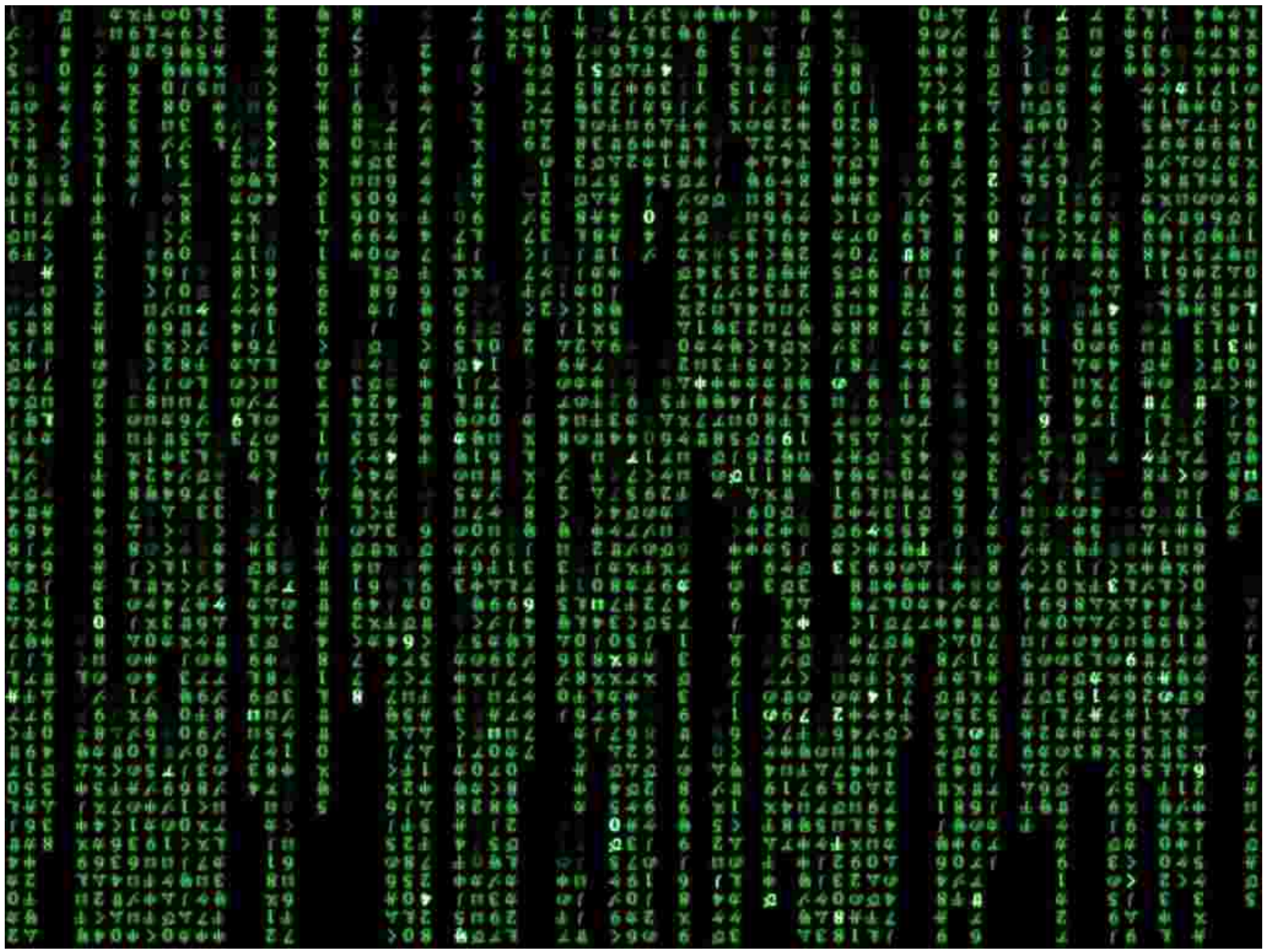
As an aside …

- How can anybody actually read these huge, expensive chunks of paper?

Sorry, back to the program …

Finding signatures of unproductive behavior

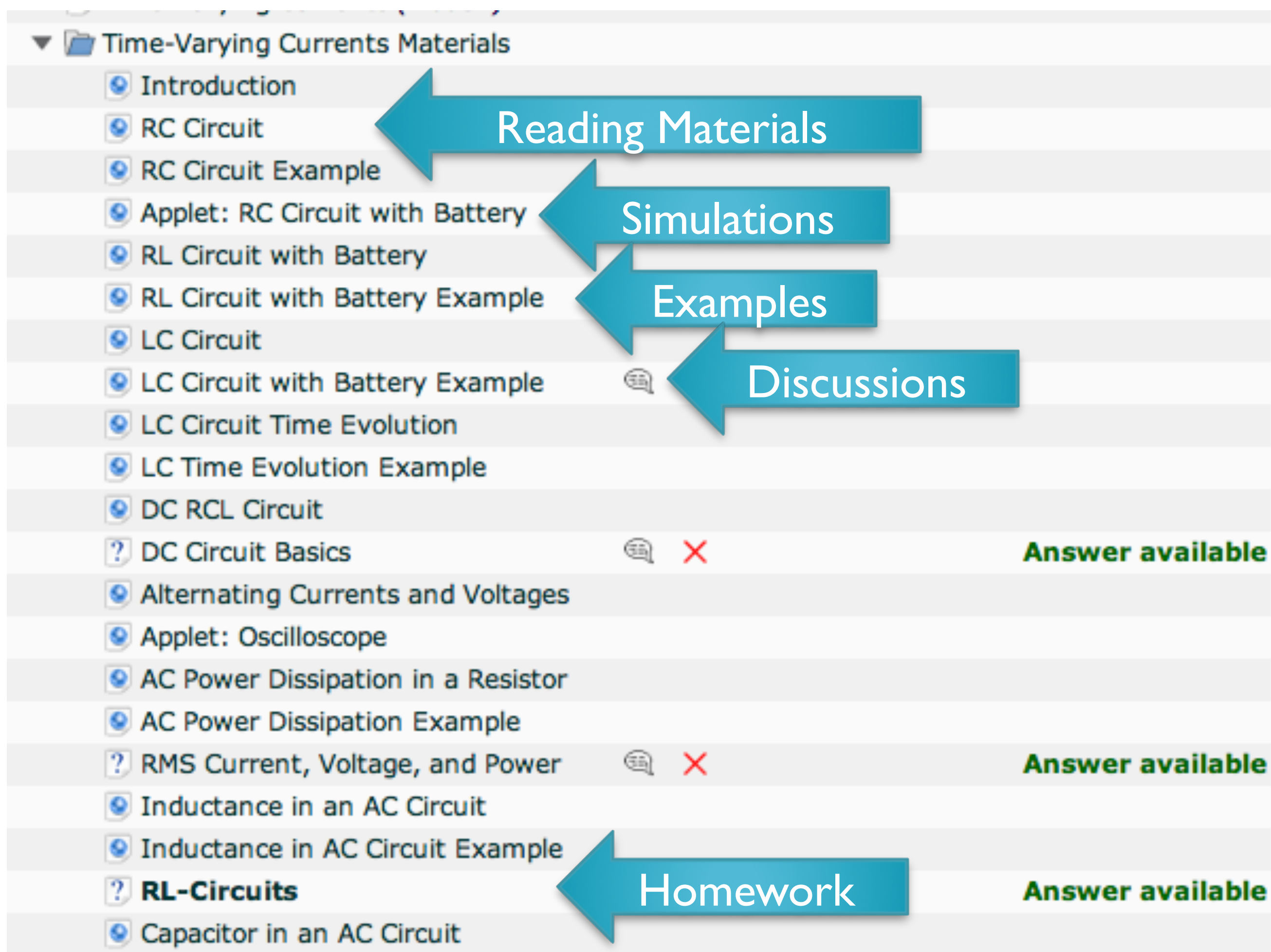
Common to all online scenarios: data!



Data Mining Access Logs

Typical online course materials

Data on materials and homework

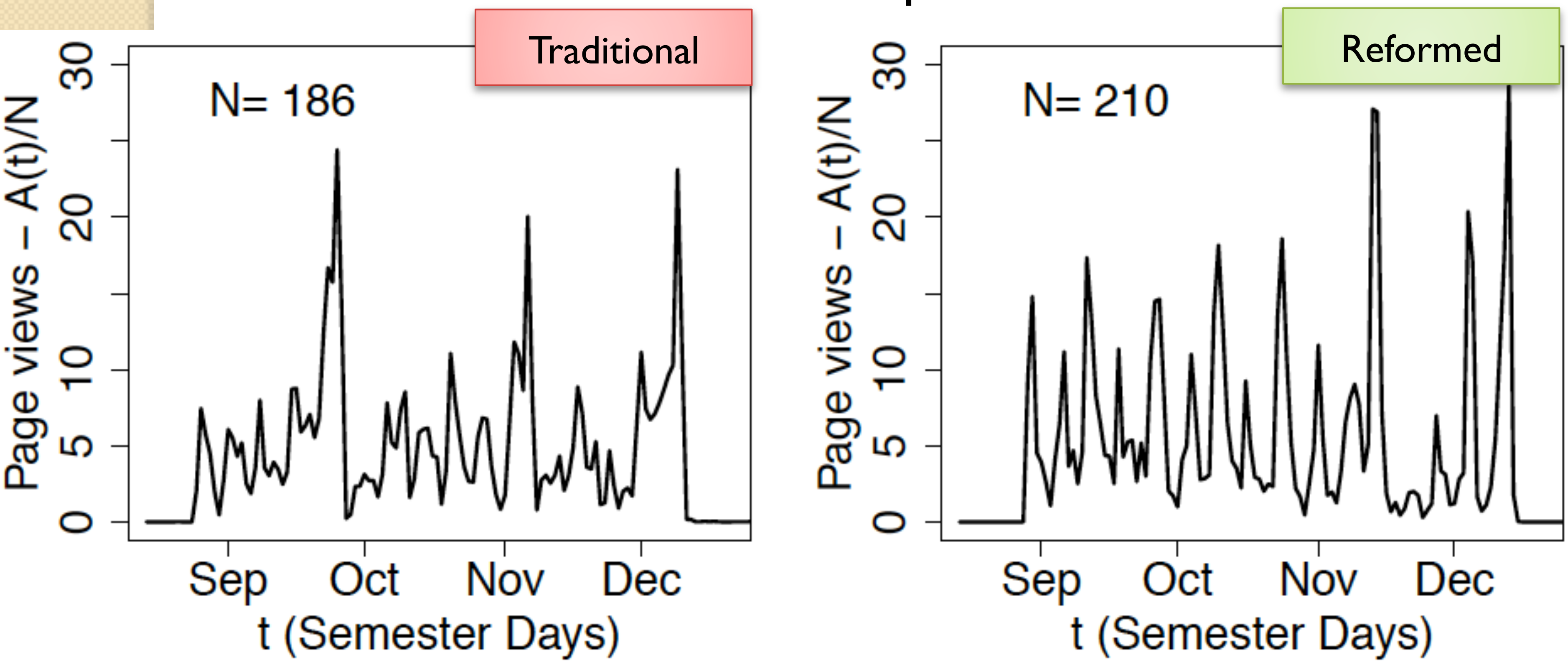


Course Structure

- Looking at different course structures:
  - Traditional course: few high-stake exams
  - Reformed course: frequent, short quizzes, peer-instruction, frequent conceptual homework
- Same online textbook materials for both

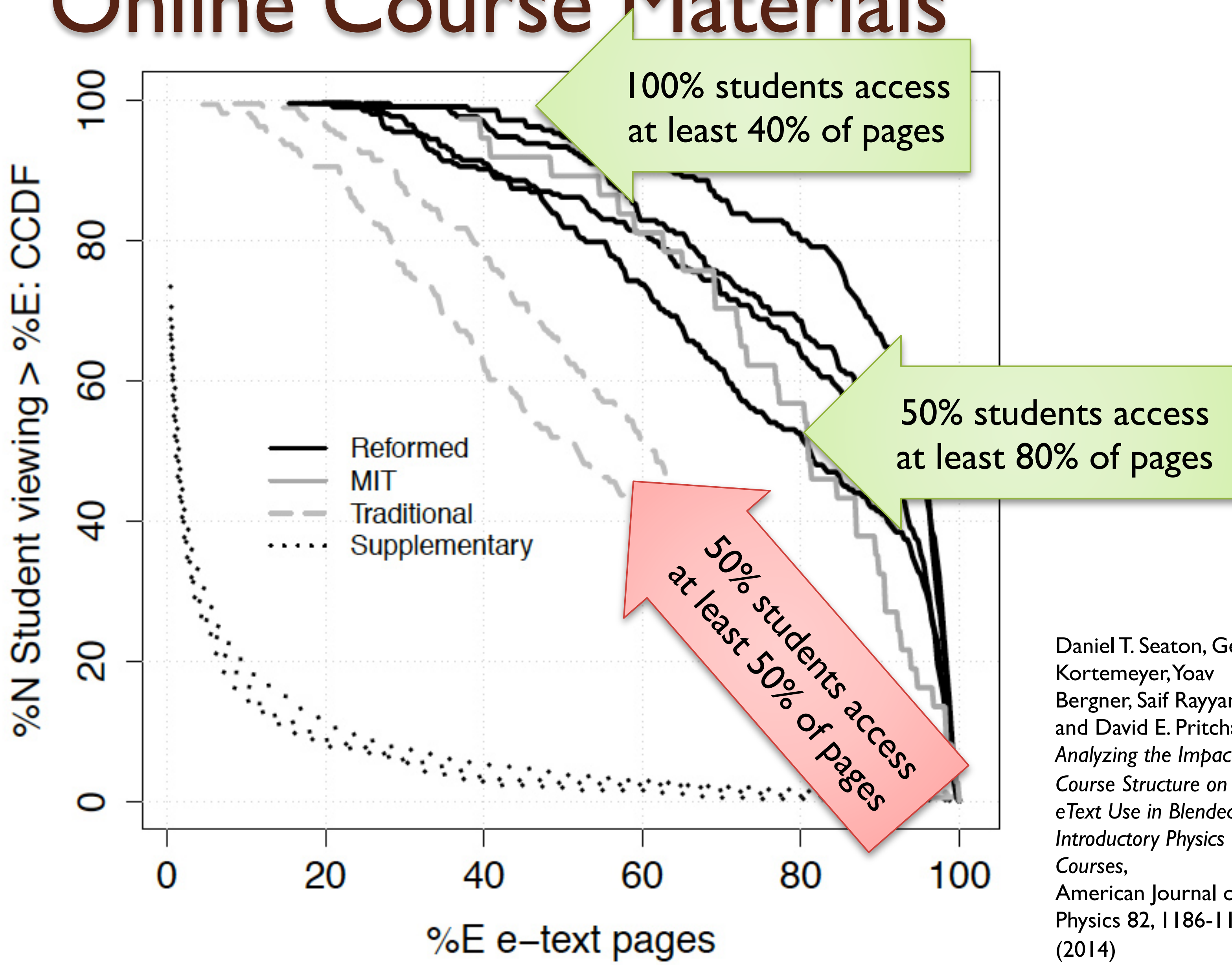
Online Course Materials

- Online course material access – cramming
- Average page views per day per student
- Guess when exams took place



Daniel T. Seaton, Gerd Kortemeyer, Yoav Bergner, Saif Rayyan, and David E. Pritchard, *Analyzing the Impact of Course Structure on eText Use in Blended Introductory Physics Courses*, American Journal of Physics 82, 1186-1197 (2014)

Online Course Materials



Daniel T. Seaton, Gerd Kortemeyer, Yoav Bergner, Saif Rayyan, and David E. Pritchard, *Analyzing the Impact of Course Structure on eText Use in Blended Introductory Physics Courses*, American Journal of Physics 82, 1186-1197 (2014)

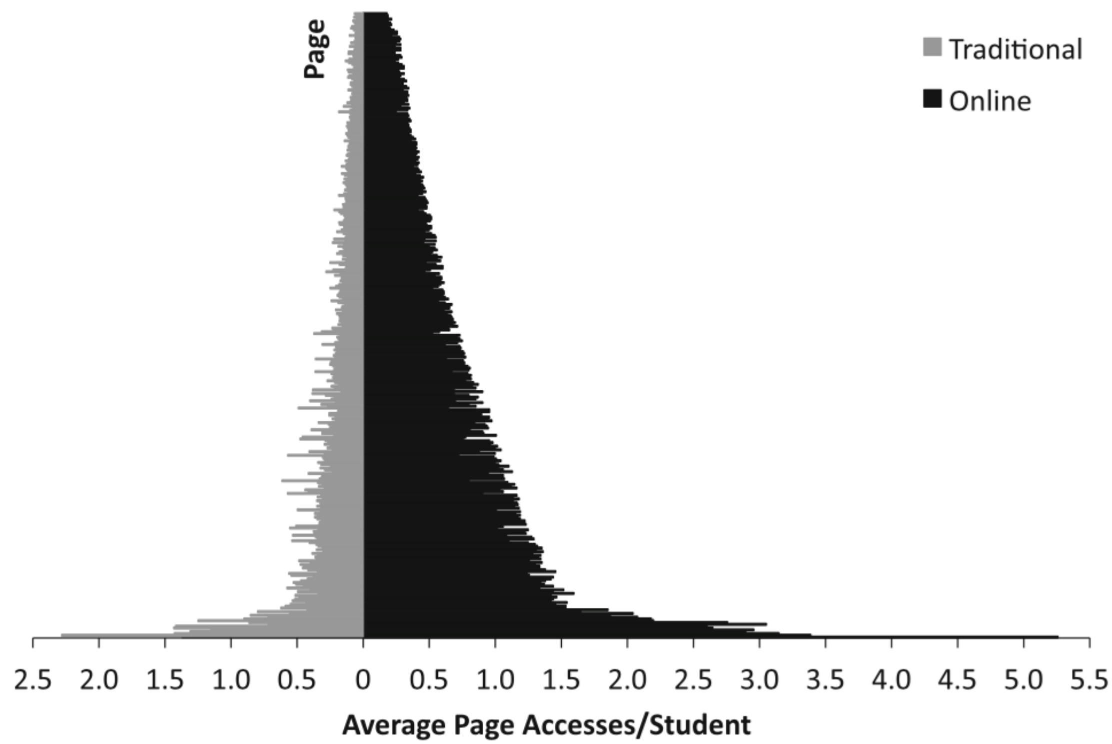
Completely Online versus Blended

- Class reform of blended courses help
- What about completely online?
- Other course:
  - One section: completely online
  - Other section: only difference that there are traditional lectures
- Everything else the same
- Students self-select

Completely Online versus Blended

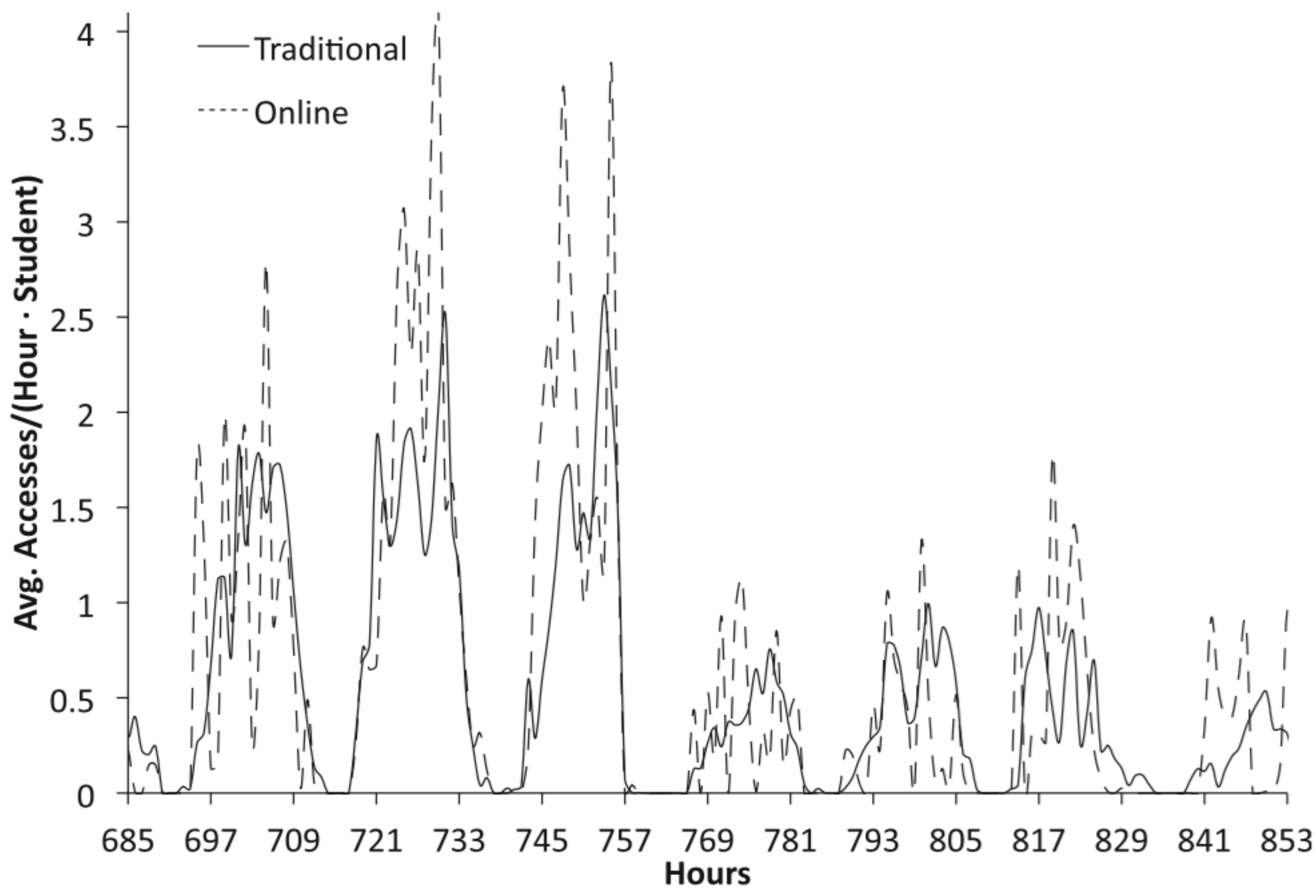
- Students in blended class read less pages than in online class
- Everybody does online homework

	Any	Pages only	Problems only
Traditional	2364	87	2276
Online	2500	265	2235



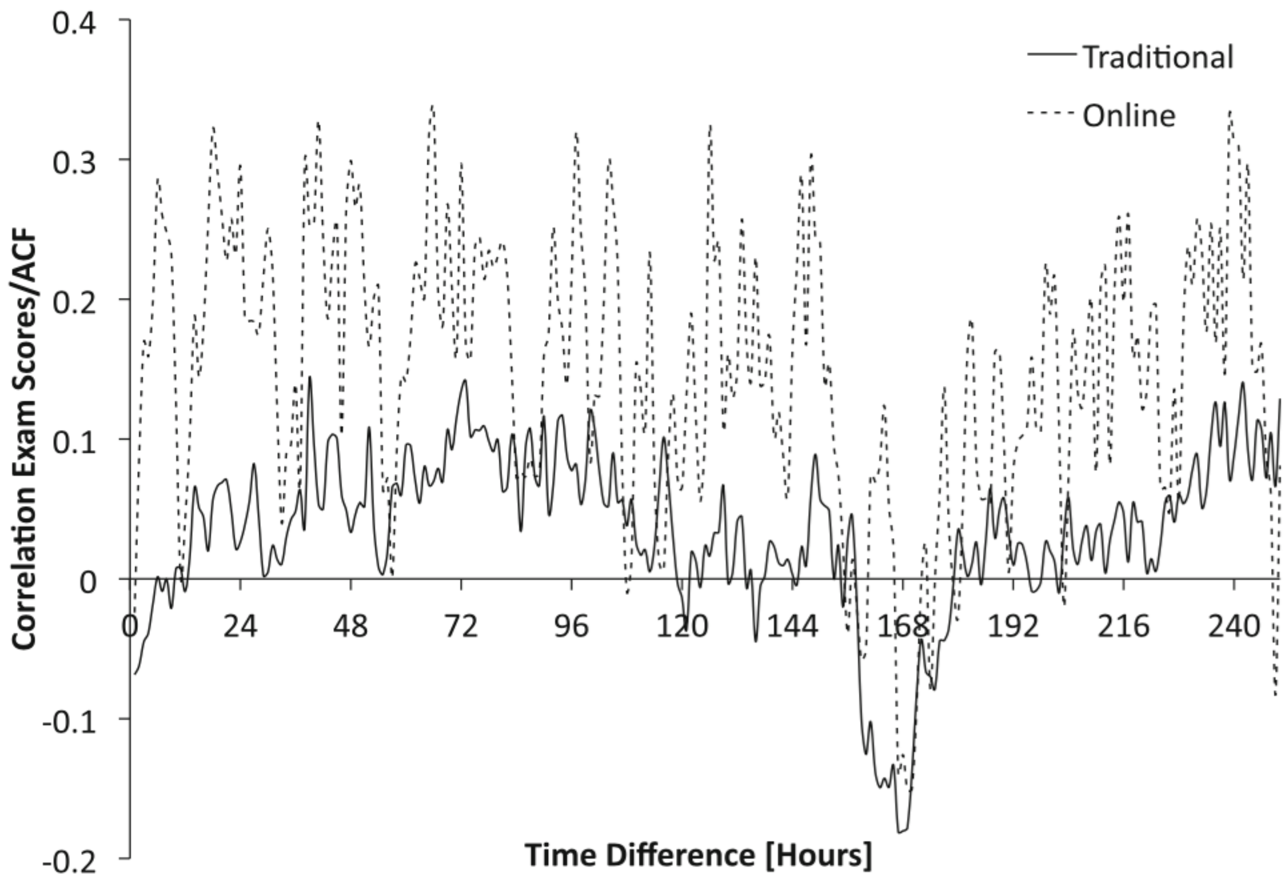
Completely Online versus Blended

- Students in online class work more irregularly
- Typical week



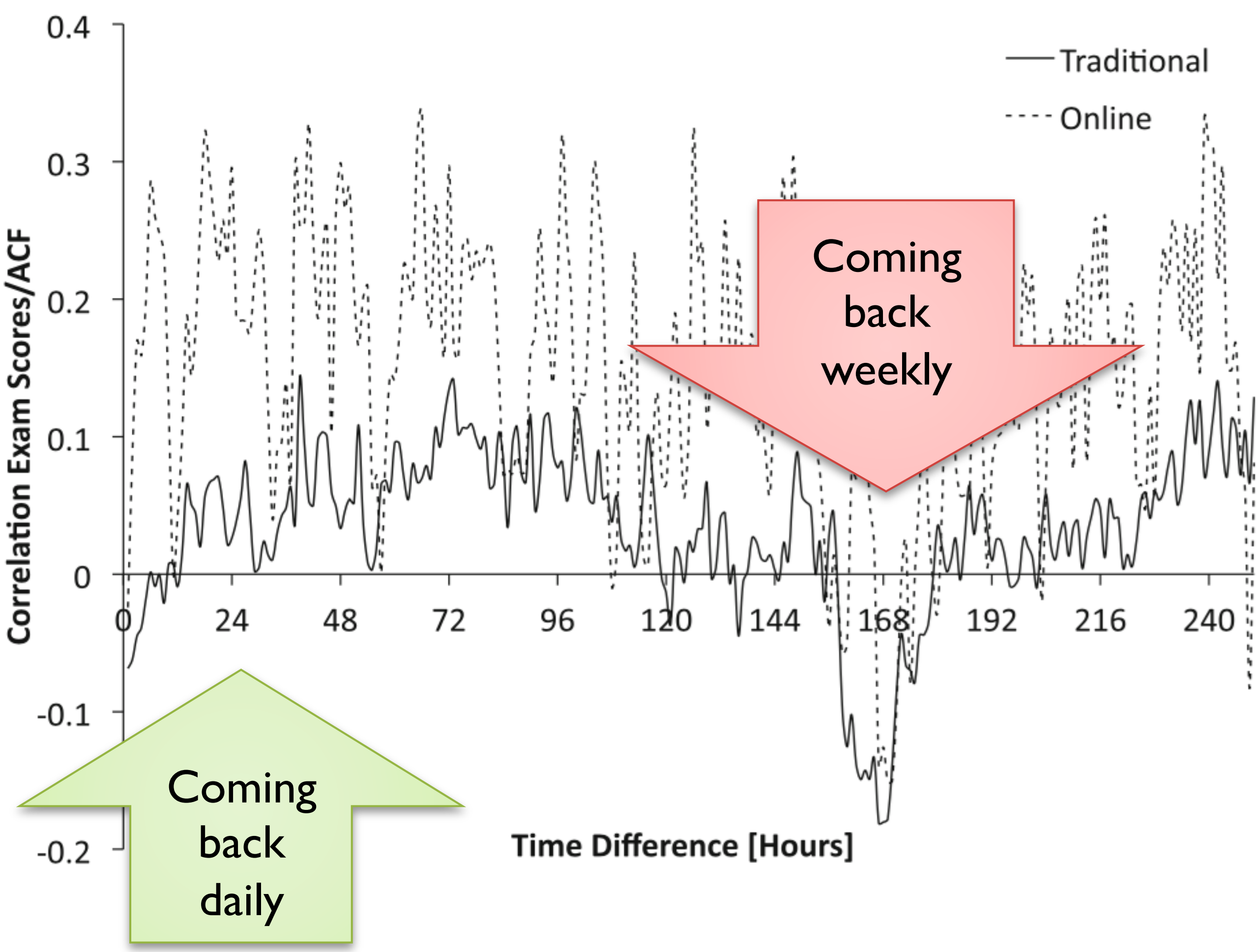
Completely Online versus Blended

- Auto-Correlation Function of Accesses versus exam scores



Completely Online versus Blended

- Auto-Correlation Function of Accesses versus exam scores



Completely Online versus Blended

- Interestingly, most significant for the students in the **blended** sections
- Problems more important than text

Type	Sections	Intercept ϵ	Hourly β_{hourly}	Daily β_{daily}	Weekly β_{weekly}
Any	Traditional	57.987***	0.009	0.425	−0.562***
Any	Online	59.374***	−0.173	2.723***	−0.279
Pages only	Traditional	55.186***	2.253*	−4.943	36.434
Pages only	Online	57.346***	−0.186	9.053*	25.153*
Problems only	Traditional	58.515***	−0.027	0.406	−0.560***
Problems only	Online	62.977***	−0.371	3.128*	−0.369

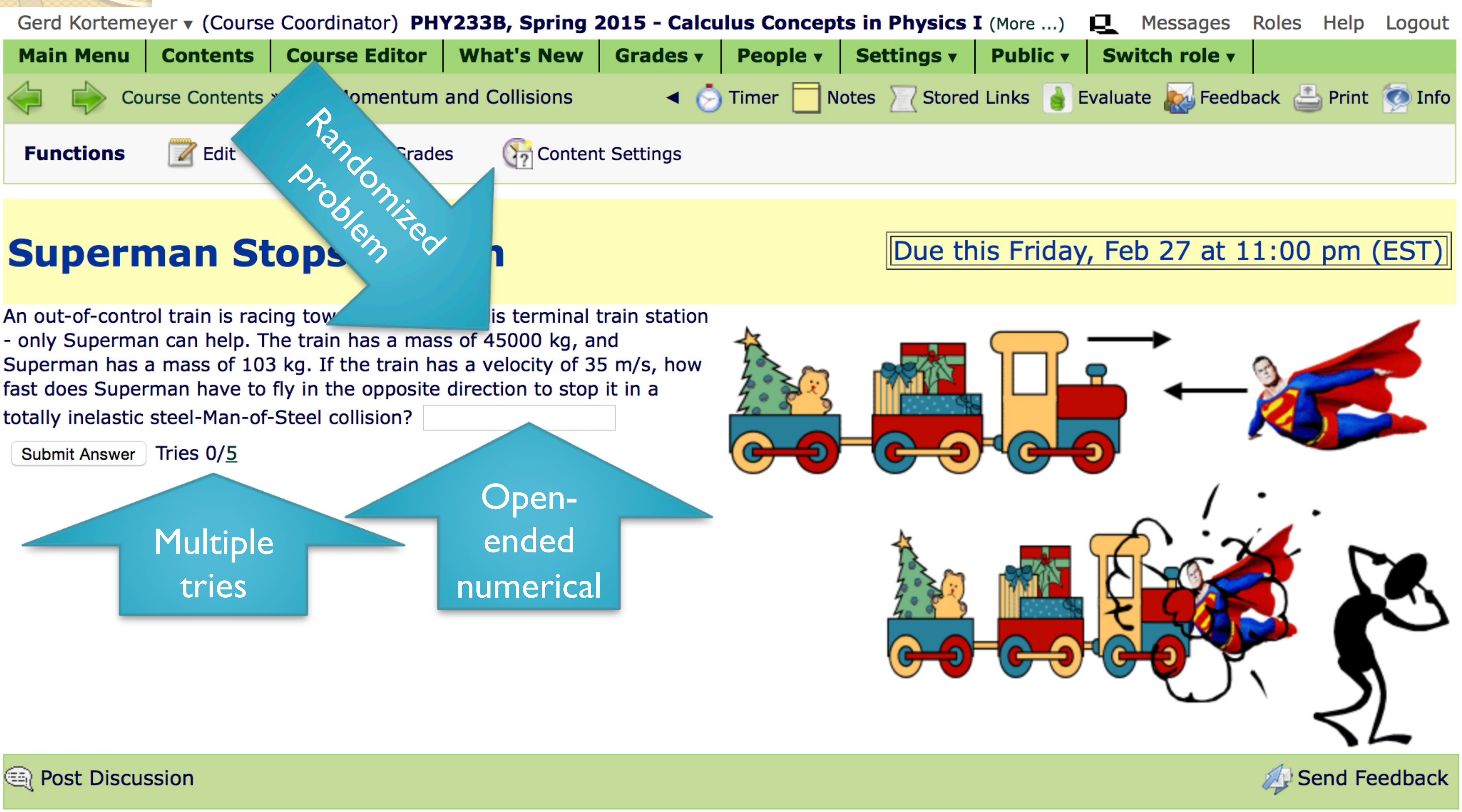
* indicates $p < 0.05$; while *** indicates $p < 0.001$

Online Course Materials

Conclusion:

- Students don’t really “read the book”
  - Unless you run a reformed course with more formative assessment
  - Nothing new …
- BUT: students do homework!
  - Let’s look at online homework

Online Homework



Online Homework

- Online behavioral features:
  - Number of tries before correct answer
  - Correct on first try
  - Total time spent on problem
  - Discussion participation
  - Working close to deadline
  - Giving up versus working up to deadline
  - First access of problem set after becoming available
  - …, etc, etc, etc, … you can define as many as you want

Online Homework

- See how well you can predict course grade from this online behavior

		Performance %		
Classifier		**2-Classes**	**3-Classes**	**9-Classes**
Tree Classifier	**C5.0**	80.3	56.8	25.6
	CART	81.5	**59.9**	**33.1**
	QUEST	80.5	57.1	20.0
	CRUISE	81.0	54.9	22.9
Non-tree Classifier	**Bayes**	76.4	48.6	23.0
	1NN	76.8	50.5	29.0
	kNN	**82.3**	50.4	28.5
	Parzen	75.0	48.1	21.5
	MLP	79.5	50.9	-
	CMC	**86.8**	**70.9**	**51.0**

Online Homework

- See how well you can predict course grade from this online behavior

Classifier		Performance %		
Classifier		2-Classes	3-Classes	9-Classes
Tree Classifier	C5.0	80.3	56.8	25.6
	CART	81.5	**59.9**	**33.1**
	QUEST	80.5	57.1	20.0
	CRUISE	81.0	54.9	22.9
Non-tree Classifier	Bayes	76.4	48.6	23.0
	1NN	76.8	50.5	29.0
	kNN	**82.3**	50.4	28.5
	Parzen	75.0	48[illegible]	21.5
	MLP	79[illegible]	5[illegible]	-
	CMC	**86.8**	**70.9**	**51.0**

Pass/Fail

High/Medium/Low

Online Homework

- Most important features

FEATURE IMPORTANCE IN 3-CLASSES USING ENTROPY CRITERION

Feature	Importance %
Total_Correct _Answers	100.00
Total_Number_of_Tries	58.61
First_Got_Correct	27.70
Time_Spent_to_Solve	24.60
Total_Time_Spent	24.47
Communication	9.21

Online Homework

- What does that mean?
  - Most important: did the student solve homework problems eventually?
  - Second: not too many tries
  - Third (factor four lower!): did they get it right on the first attempt?
- Tenacity more important than immediate genius!

B. Minaei-Bidgoli, D.A. Kashy, G. Kortemeyer, and W. Punch,
Predicting Student Performance: an Application of Data Mining Methods with an Educational Web-Based System (LON-CAPA),
Frontiers in Education Conference 2003

Online Homework

- What does that mean?
  - Most important: did the student solve homework problems eventually?
  - Second: not too many tries
  - Third (factor four lower!): did they get it right on the first attempt?

All related to number of tries

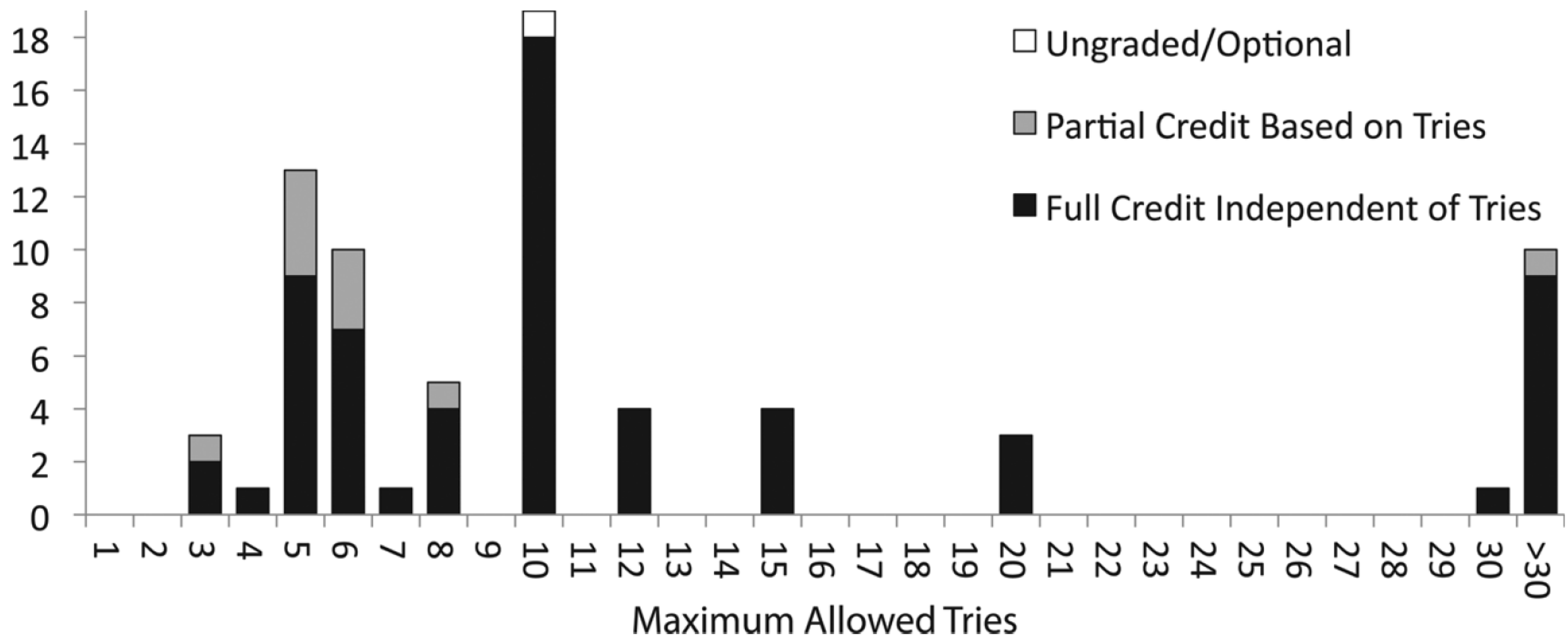
- Tenacity more important than immediate genius!

B. Minaei-Bidgoli, D.A. Kashy, G. Kortemeyer, and W. Punch,
Predicting Student Performance: an Application of Data Mining Methods with an Educational Web-Based System (LON-CAPA),
Frontiers in Education Conference 2003

Typical Online Physics Problem

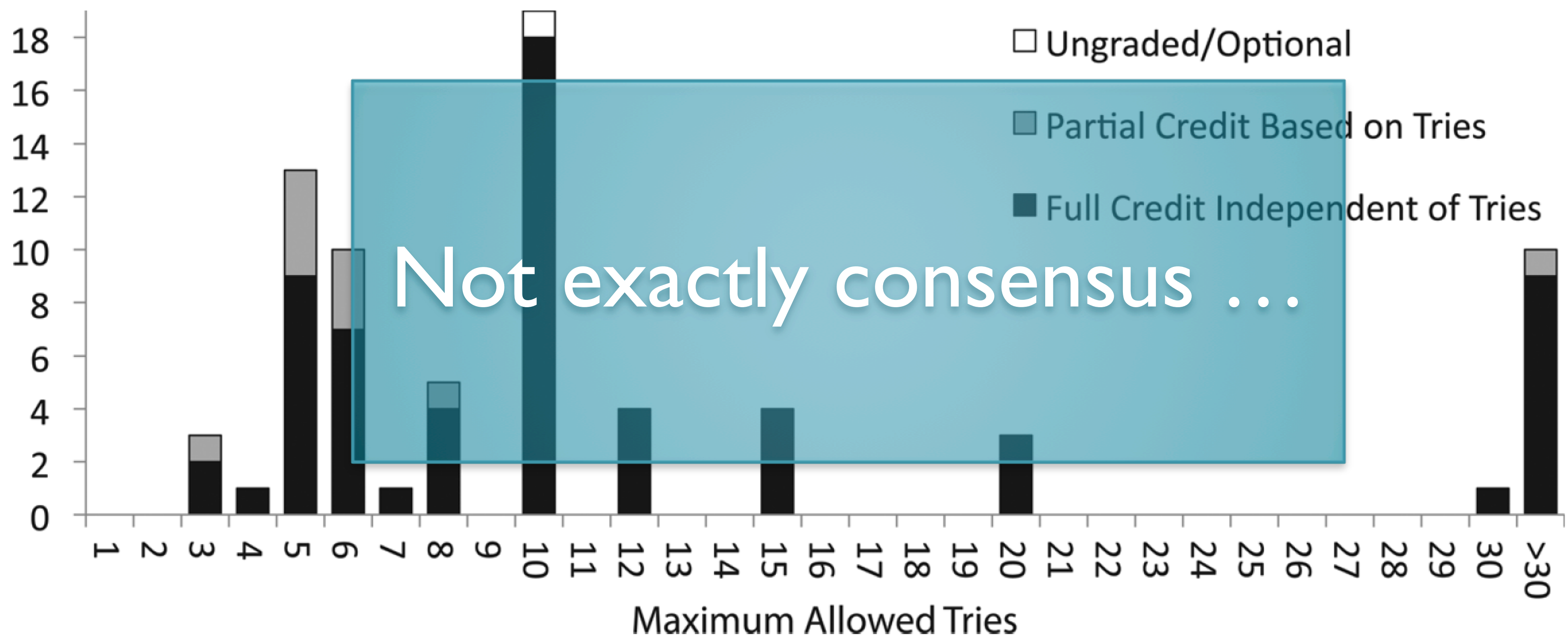
How Many Tries to Grant?

- Quick survey among 74 PER faculty and LON-CAPA users
- Self-identified as instructors-of-record



How Many Tries to Grant?

- Quick survey among 74 PER faculty and LON-CAPA users
- Self-identified as instructors-of-record

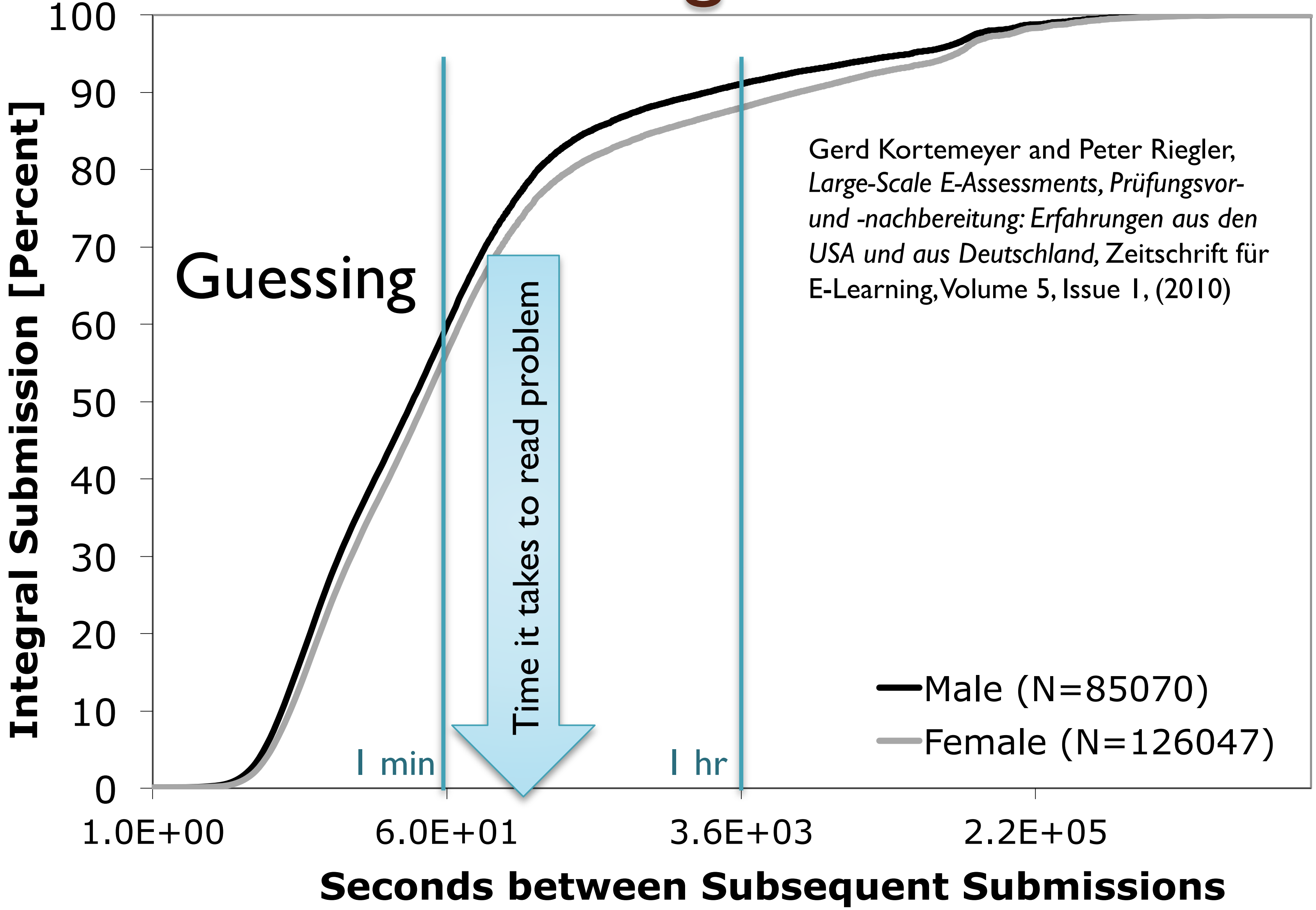


How Many Tries to Grant?

- Why is there no consensus?
- Balancing act

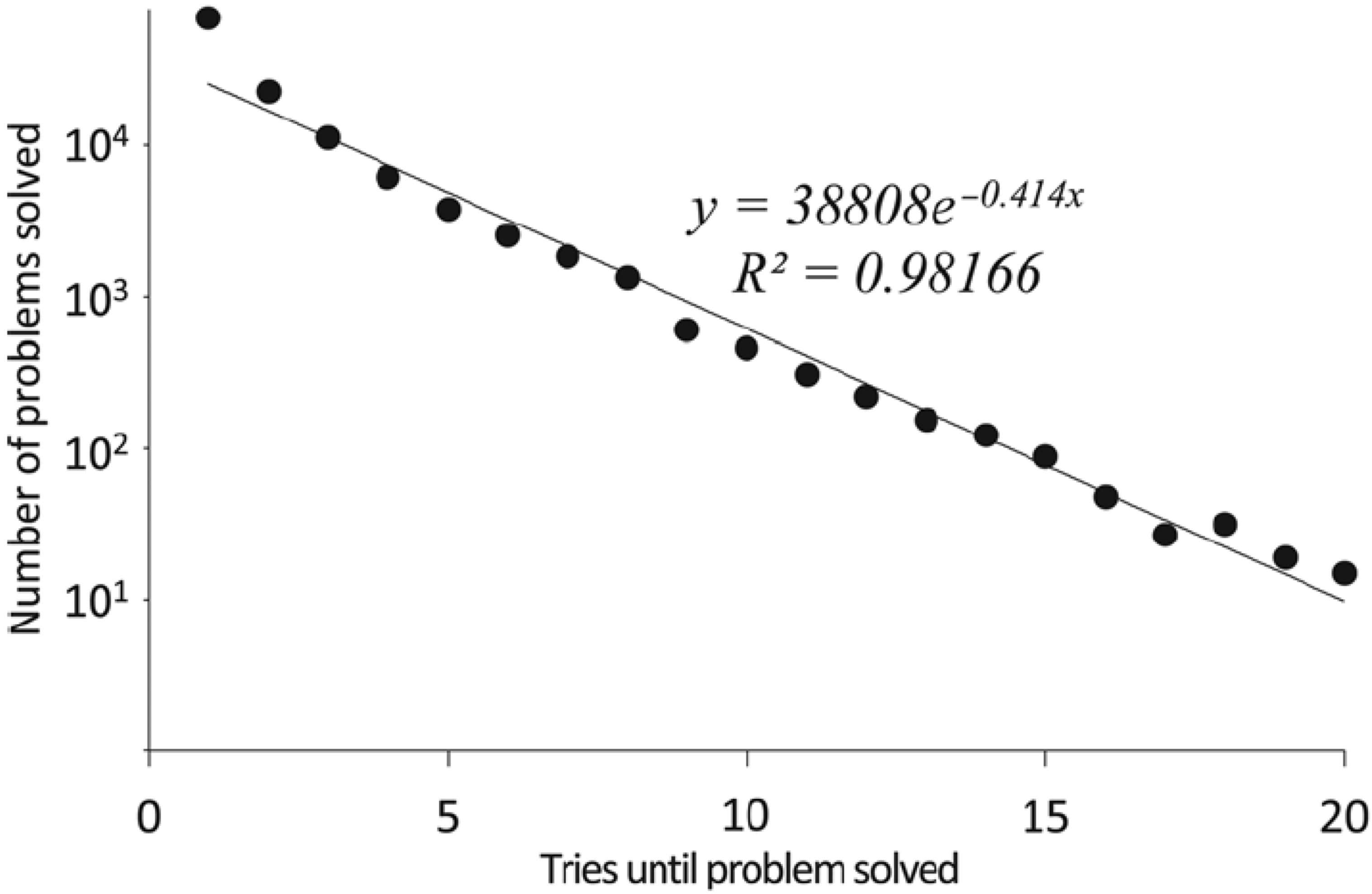
	Low Number of Allowed Tries	High Number of Allowed Tries
Possibly Good	• Better exam preparation • Less grade-inflation	• Better mastery-based formative assessment • Encouragement • Less whining
Possibly Bad	• Discouragement • Copying • More whining	• Random guessing • False sense of security

Random Guessing



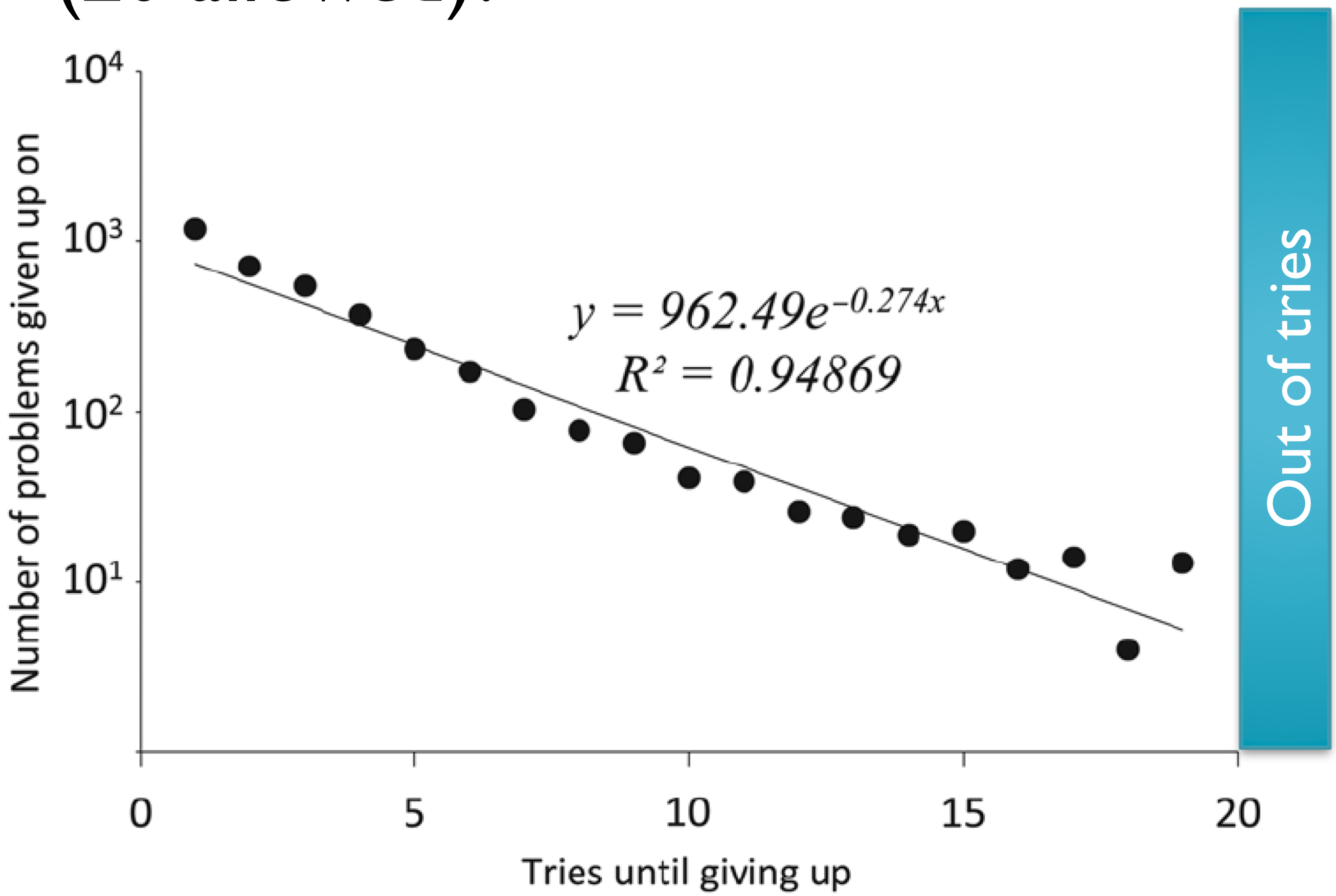
Tries versus Success

- How many tries does it take (20 allowed)?



Tries versus Giving Up

- After how many tries do students give up (20 allowed)?



Tries Follow Decay Laws!

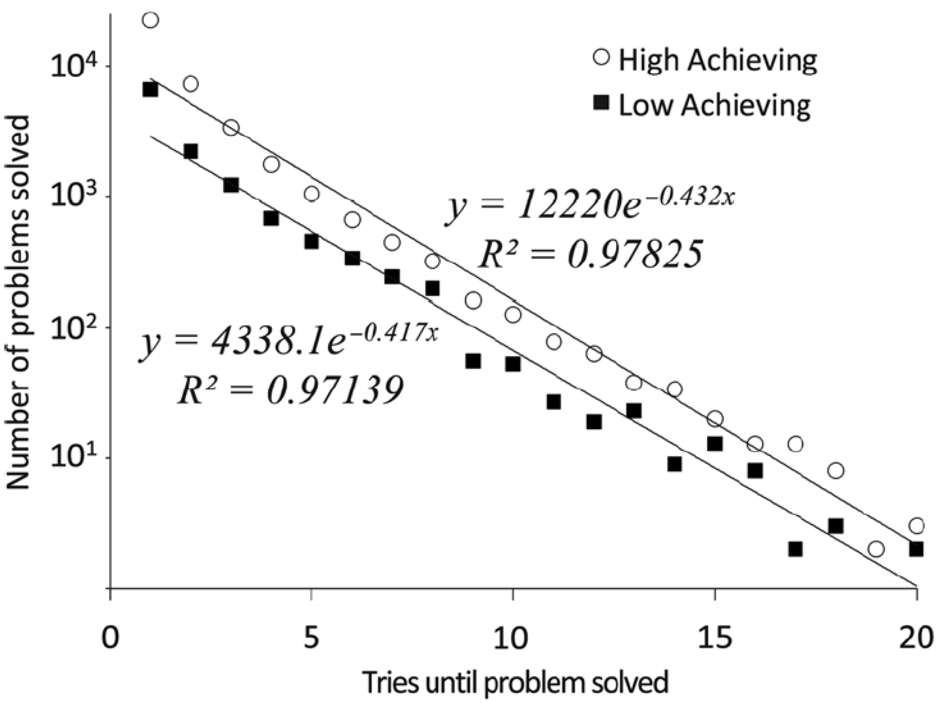
- Comparing three classes:
  10 tries, 12 tries, and 20 tries max.
- Surprisingly**, for all these classes**, both success and giving up follow

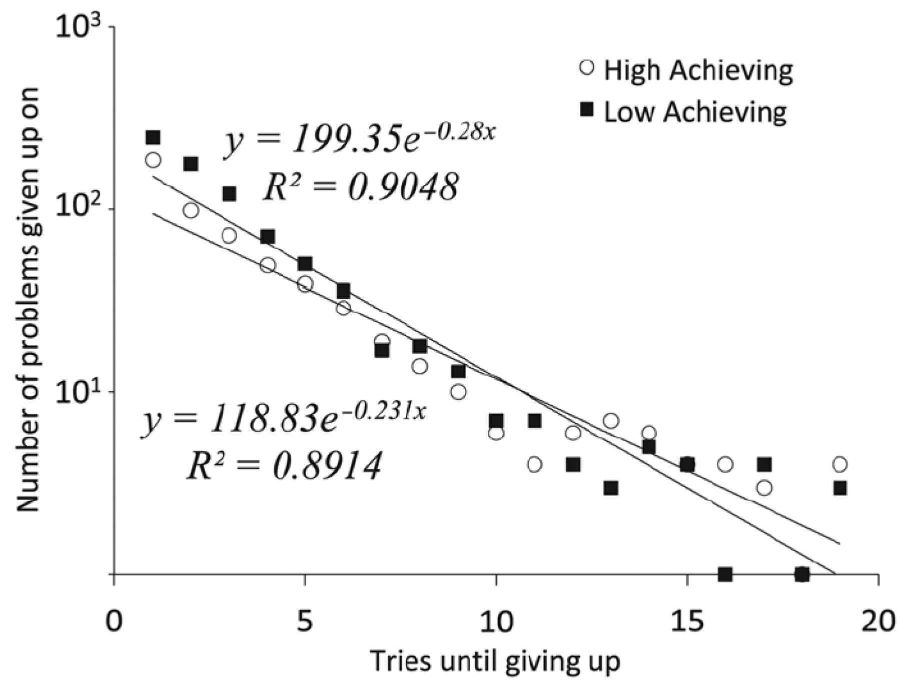
$$
\begin{aligned}
\Delta N_s(n) &= N_{s,0} \exp(-\lambda_s n) \\
\Delta N_a(n) &= N_{a,0} \exp(-\lambda_a n)
\end{aligned}
$$

- Tries are independent of each other!
- Lambdas are like probabilities
- **Students do not learn from their previous mistakes!**

Tries versus Success

- Is it just the low-achieving students who do not learn from previous failures?

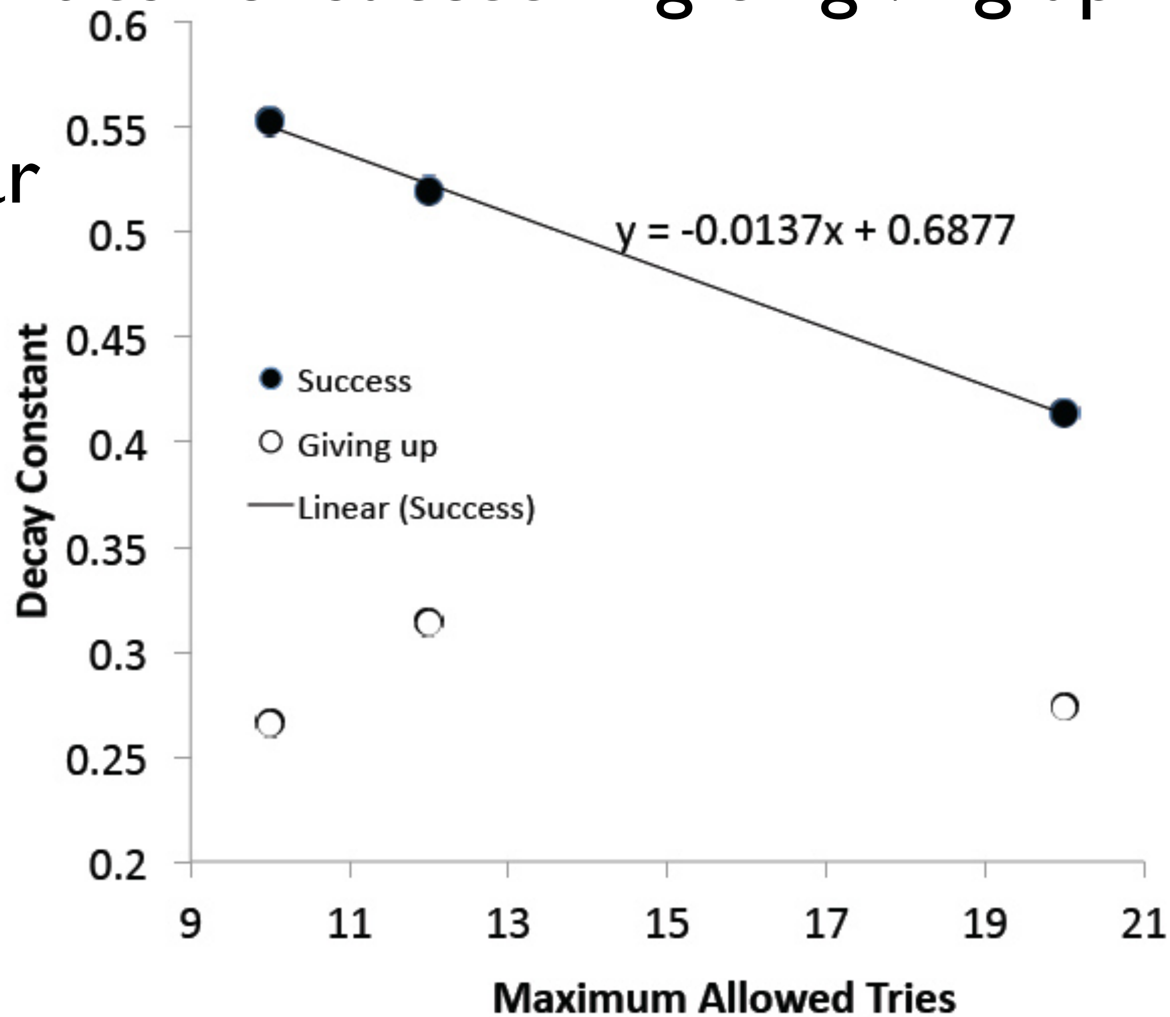




No.

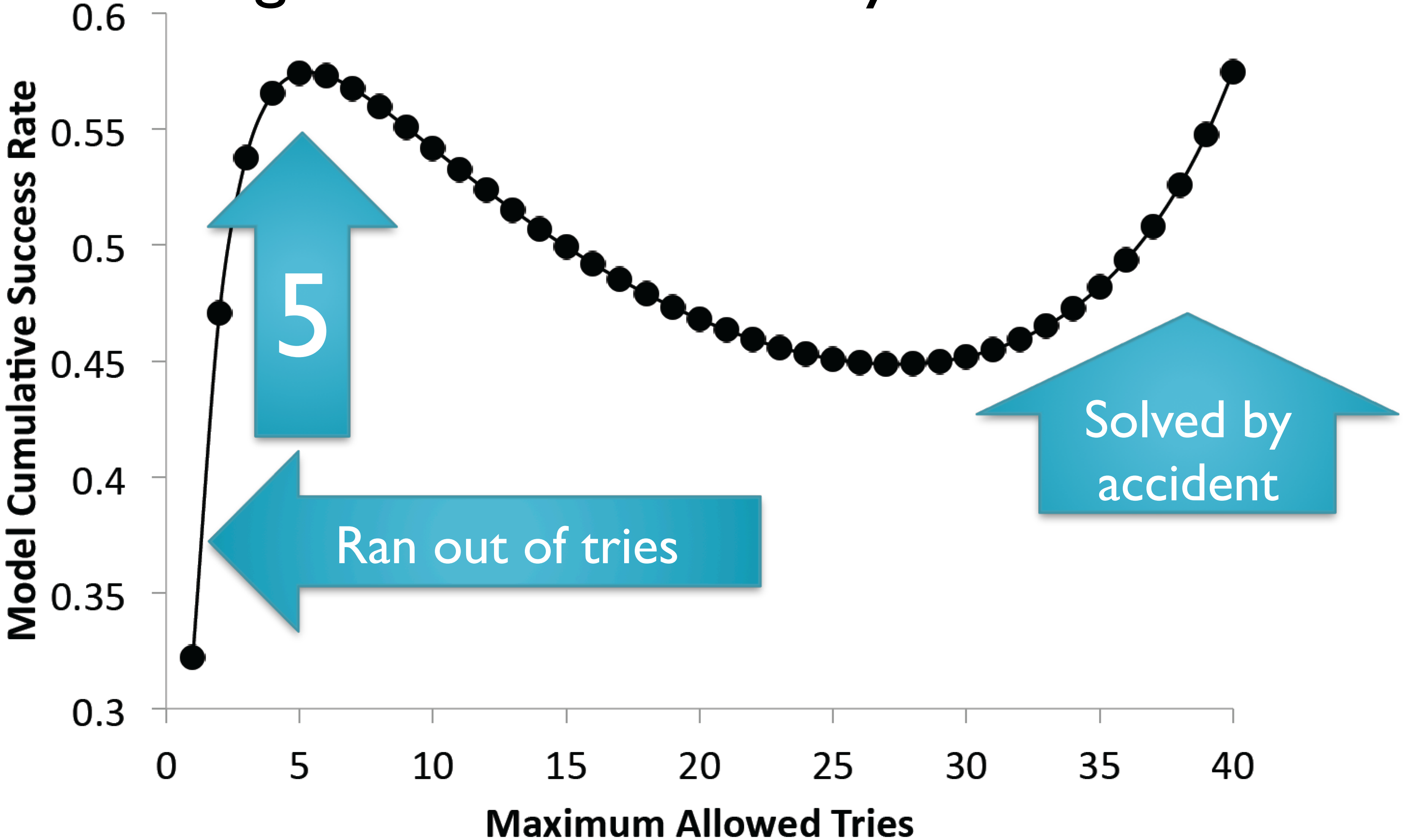
Tries versus Success

- "Probabilities" of succeeding or giving up on a particular attempt



Tries versus Success

- Using this model of "decay constants"

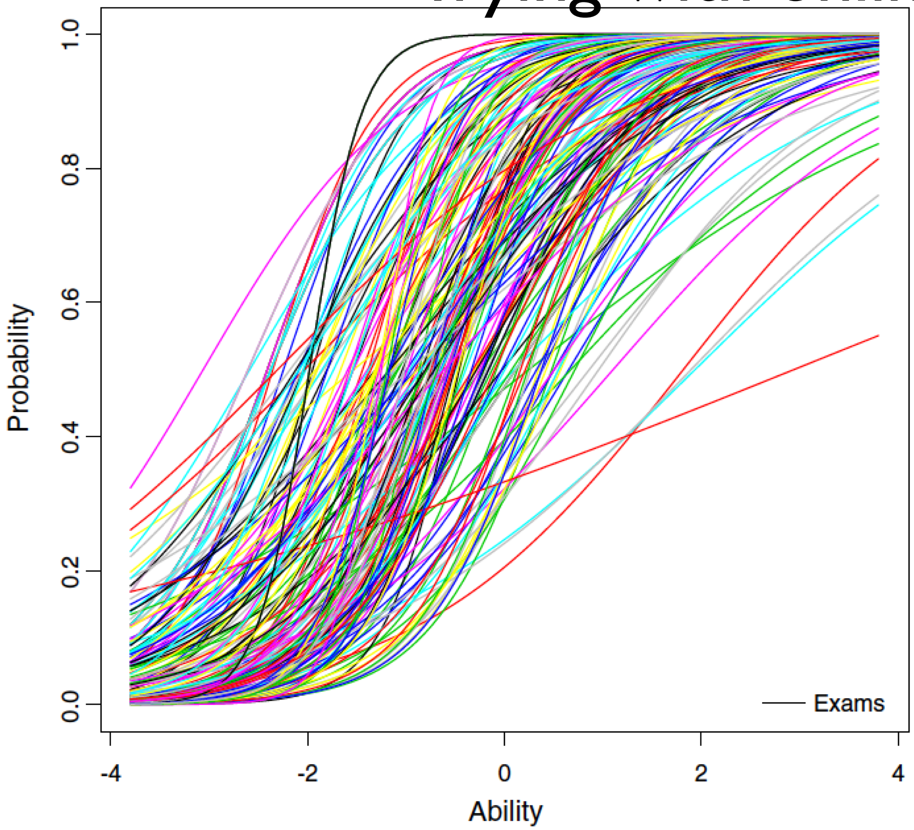


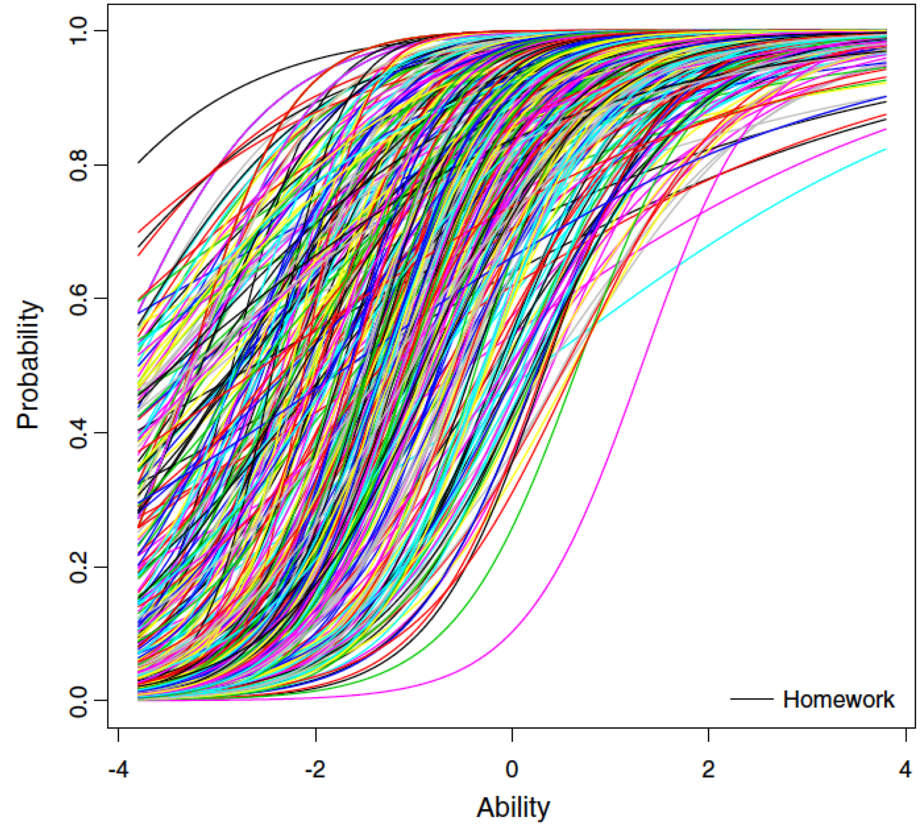
Hmm …

- A lot depends on homework
- How meaningful is online homework?

Item Response Theory

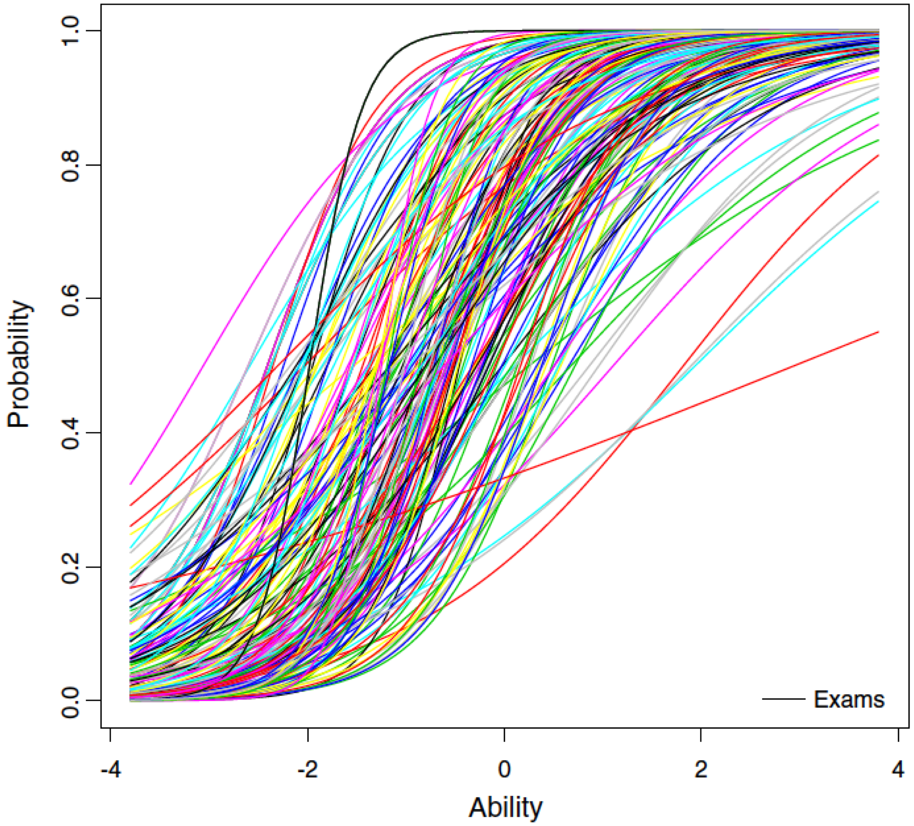
- IRT was developed for summative assessments
  - Trying with online homework

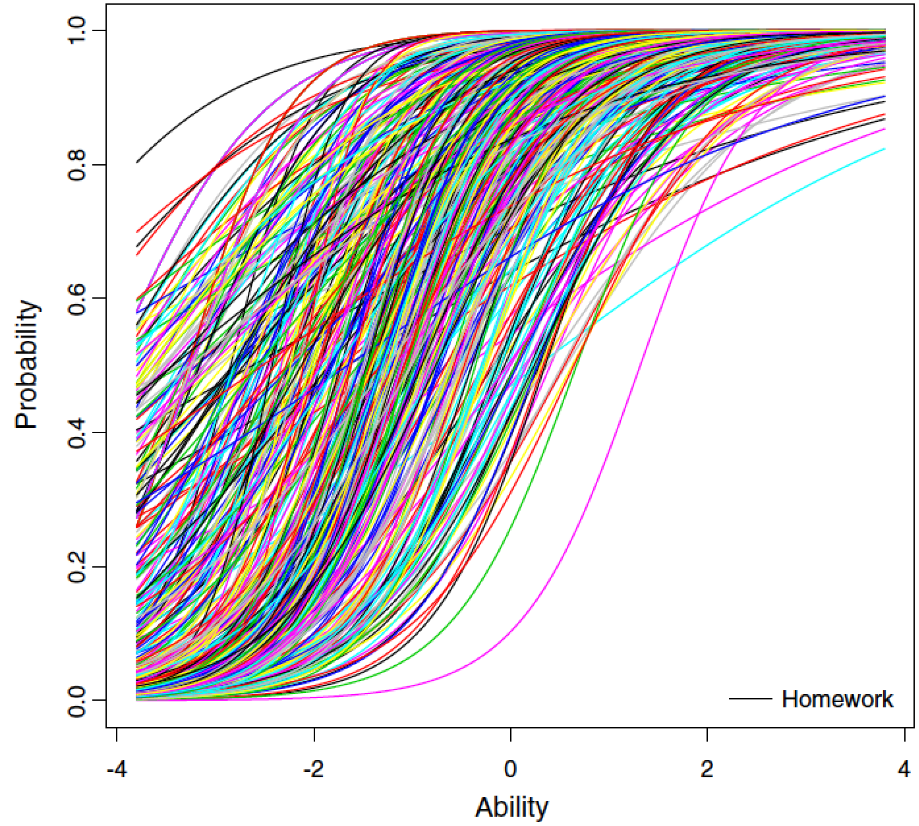




Item Response Theory

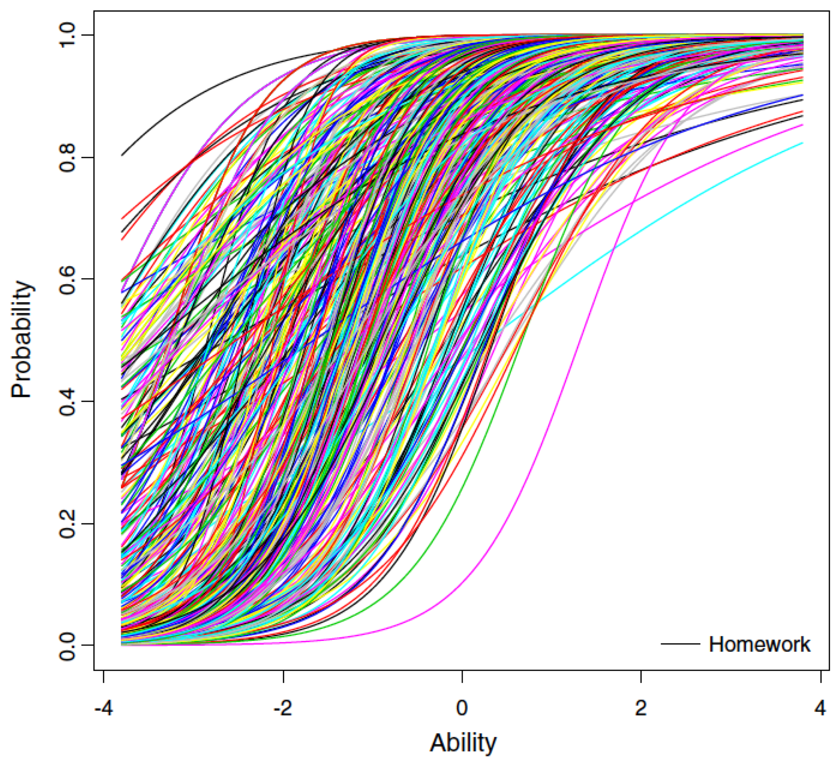
- You can see the "noise"
- This is guessing and copying

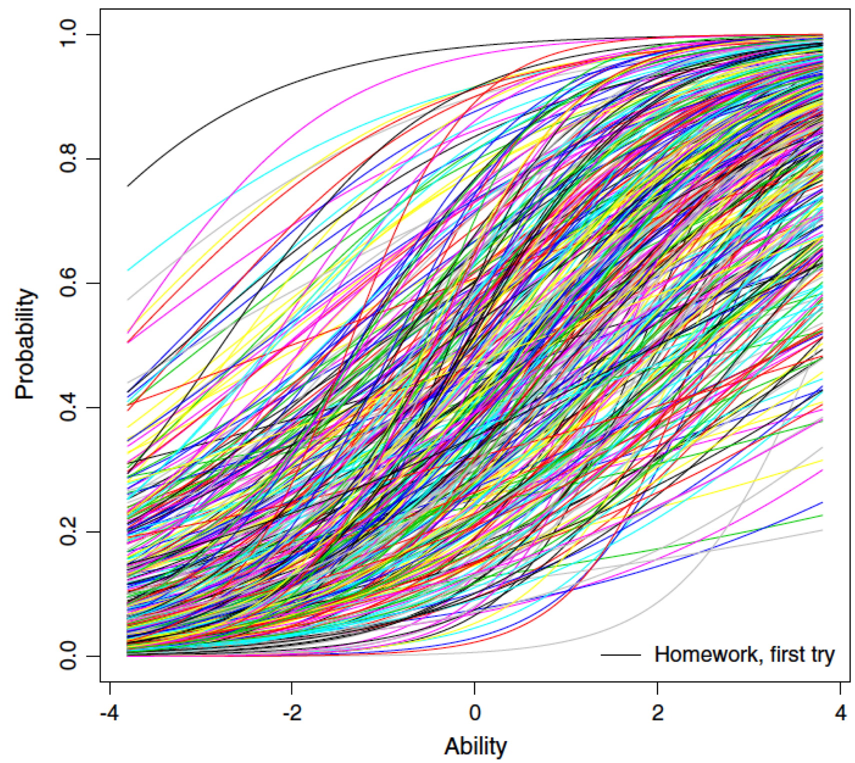




Item Response Theory

- Having finished homework eventually is more meaningful than on the first try
  - We already knew that …





Item Response Theory

- IRT can be used for online homework
- Final result ability better predictor of exam ability
- However, best predictor:
  first try during the first quarter of the semester!
  - Unproductive behavior increases over the course of the semester!

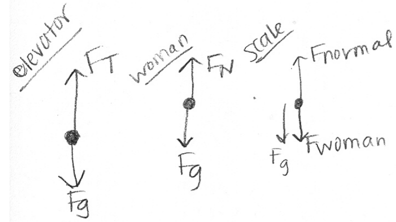
Gerd Kortemeyer, *Extending Item Response Theory to Online Homework*, Phys. Rev. ST Phys. Educ. Res. 10, 010118 (2014)

Why?

- Why do students not learn from their previous failed attempts?
- By being able to try again, they should have a chance to verify their solutions and think through the physics.
- Why is this opportunity apparently wasted?

Why?

- Prime suspect: plug-and-chug
- Just plugging numbers from one equation into the next
- No chance to backtrack
- No chance to do dimensional analysis, etc., etc.



a) $F = ma$

$\Sigma F_{y\,elevator} = F_T - F_g = ma$

$9410 - (815+60)9.8 = (815+60)(a)$

$\frac{835}{875} = a = .95\ m/s^2$ upwards

$\Sigma F_{y\,scale} = F_N - F_{g\,earth \to woman} = (60\ kg)(.95\ m/s^2)$

$F_N - 60(9.8) = 57$

$F_N = 645\ N$

If the scale is pushing up w/ 645 N, that means it has a force applied of $\boxed{645\ N}$

b) $9410 - 875(\frac{9.8}{5}) = 875a$

$a = 8.79\ m/s^2$

$F_N - F_g = 60(8.79)$

$F_N - 60(\frac{9.8}{5}) = 60(8.79)$

$F_N - 117.6 = 527.4$

$\boxed{F_N = 645\ N}$

c) $9410 - (815+12)(9.8) = (815+12)a$

$9410 - 827(9.8) = 827a$

$a = 1.58\ m/s^2$

$F_N - 12(9.8) = 12(1.58)$

$F_N - 117.6 = 18.96$

$\boxed{F_N = 136.6\ N}$

Why?

- Plug-and-chug is typical for numerical problems
- As soon as numbers appear in the problem, they apparently have to be used asap.

a) $T = (m+M)(g+a)$ scale reads: $F = m\left(\frac{T}{m+M}\right)$

$g + a = \frac{T}{m+M}$

$a = \frac{T}{m+M} - g$

b) $F = 60.0\left(\frac{9410}{60+815}\right)$

$= 645\ N$

c) $T = (m+M)\left(\frac{1}{5}g + a\right)$ $F = m\left(\frac{T}{m+M}\right)$

$a = \frac{T}{m+M} - \frac{1}{5}g$ $= 60.0\left(\frac{9410}{60+815}\right)$

$= 645\ N$

d) $F = m\left(\frac{T}{m+M}\right)$

$= 12.0\left(\frac{9410}{12+815}\right)$

$= 136.5\ N$

Why?

Gerd Kortemeyer ▾ (Course Coordinator) **PHY233B, Spring 2015 - Calculus Concepts in Physics I** (More ...) Messages Roles Help Logout

Main Menu | **Contents** | **Course Editor** | **What's New** | **Grades ▾** | **People ▾** | **Settings ▾** | **Public ▾** | **Switch role ▾**

Course Contents » ... » Momentum and Collisions ◄ Timer Notes Stored Links Evaluate Feedback Print Info

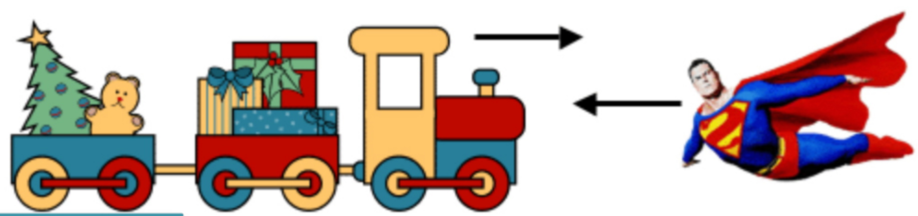
Functions Edit Content Grades Content Settings

Superman Stops a Train

Due this Friday, Feb 27 at 11:00 pm (EST)

An out-of-control train is racing toward the Metropolis terminal train station - only Superman can help. The train has a mass of 45000 kg, and Superman has a mass of 103 kg. If the train has a velocity of 35 m/s, how fast does Superman have to fly in the opposite direction to stop it in a totally inelastic steel-Man-of-Steel collision?

Submit Answer Tries 0/5

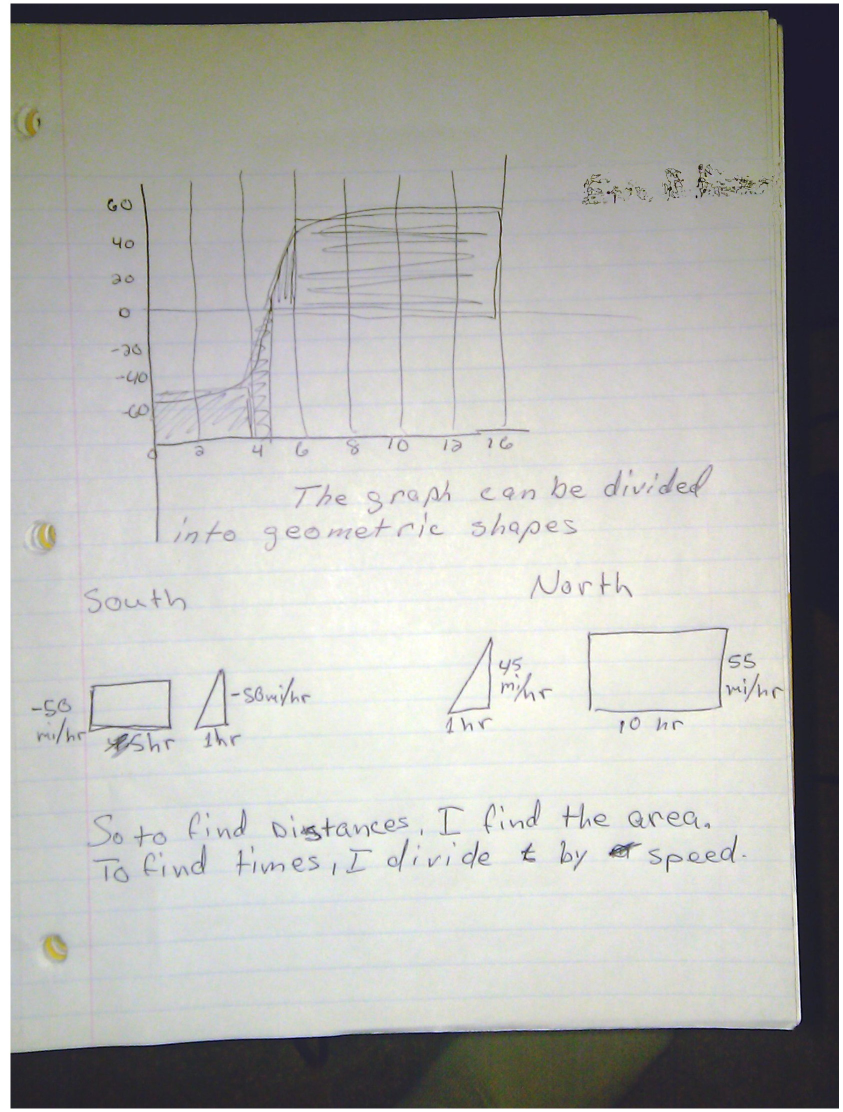


Really, these problems are not very good.
Take a bunch of numbers, plug them into equations, get another number.
Who really cares about these numbers?
What do the students really learn?

Send Feedback

Another Approach

- Curb plug-and-chug
- Have students turn in some derivations and graphs simply by photographing them with their cell phones and uploading them to the CMS



… or maybe …

- Give better homework
- Multiple-part,
  non-numeric (symbolic/conceptual),
  dynamic, randomizing scenarios
  - Less success by random guessing
    - Random guessing leads students down a garden path
  - Less chances of success by blind copying
    - Every scenario and path different
    - Students can and should discuss the physics, not just the result

… or maybe …

A plate capacitor has been charged. Its plates are then pushed closer together after they had been disconnected from the voltage source.

- ◯ The capacitance increases.
- ◯ The capacitance stays the same.
- ◯ The capacitance decreases.

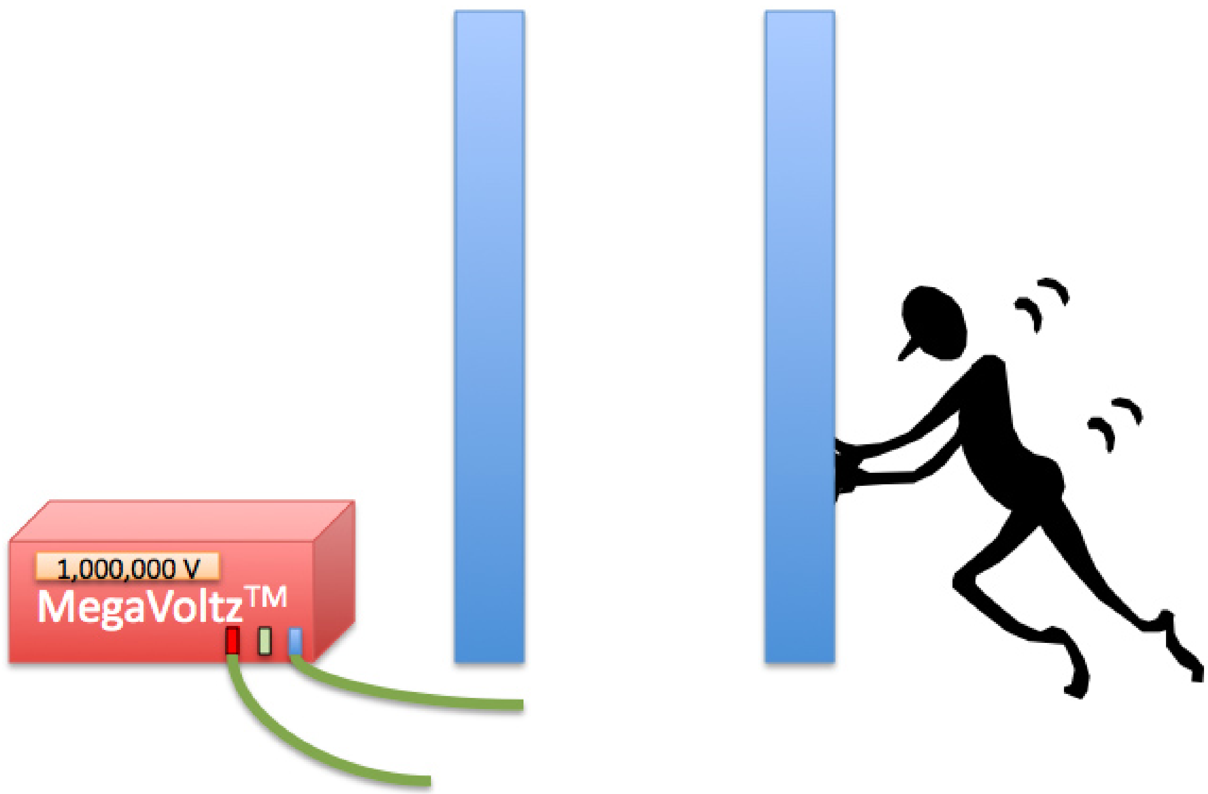
Submit Answer Tries 0

- ◯ The voltage increases.
- ◯ The voltage stays the same.
- ◯ The voltage decreases.

Submit Answer Tries 0

- ◯ The charge increases.
- ◯ The charge stays the same.
- ◯ The charge decreases.

Submit Answer Tries 0



… or maybe …

A plate capacitor has been charged. Its plates are then pulled further apart while still connected to the voltage source.

- ◯ The capacitance increases.
- ◯ The capacitance stays the same.
- ◯ The capacitance decreases.

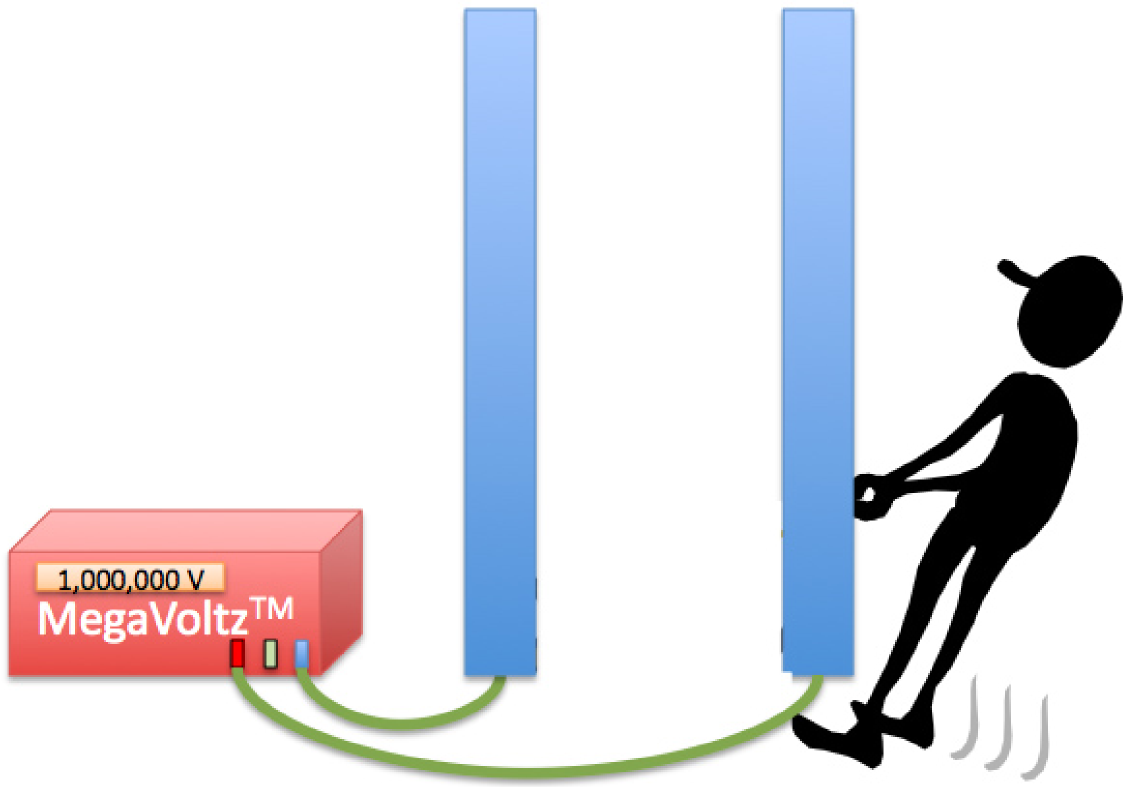
Submit Answer Tries 0

- ◯ The voltage increases.
- ◯ The voltage stays the same.
- ◯ The voltage decreases.

Submit Answer Tries 0

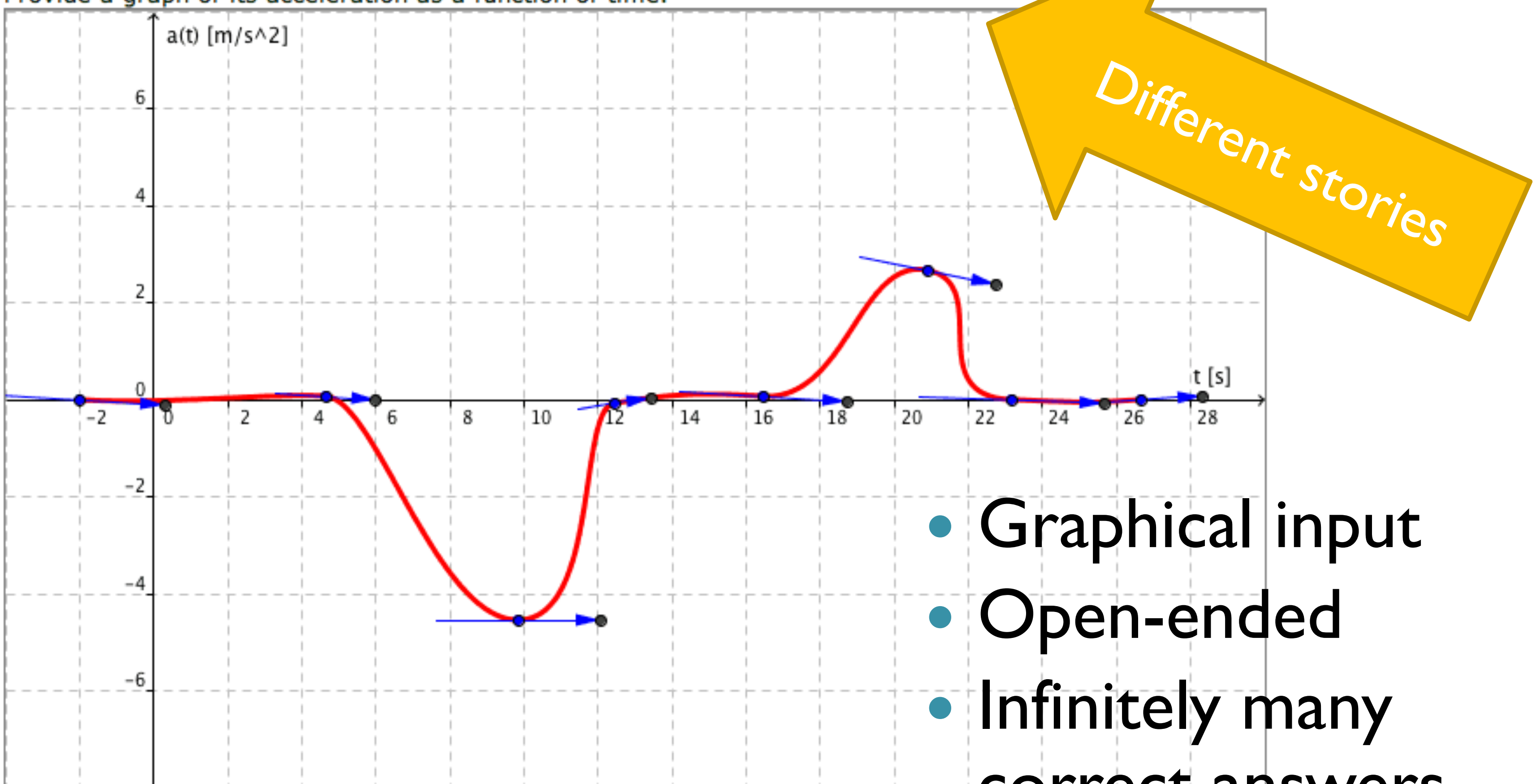
- ◯ The charge increases.
- ◯ The charge stays the same.
- ◯ The charge decreases.

Submit Answer Tries 0



… or maybe …

At t=0 s, a car cruises at a constant positive velocity. Suddenly, a light switches to red. At t=10 s, the driver is maximum on the brake. The car then stops in front of the red light for over 2 seconds. Eventually, it drives off, and then again cruises at a constant velocity. The car cannot accelerate with more than 3 m/s^2.
Provide a graph of its acceleration as a function of time.



You are correct. Computer's answer now shown above. Previous Tries

Script Vars
Rules Log

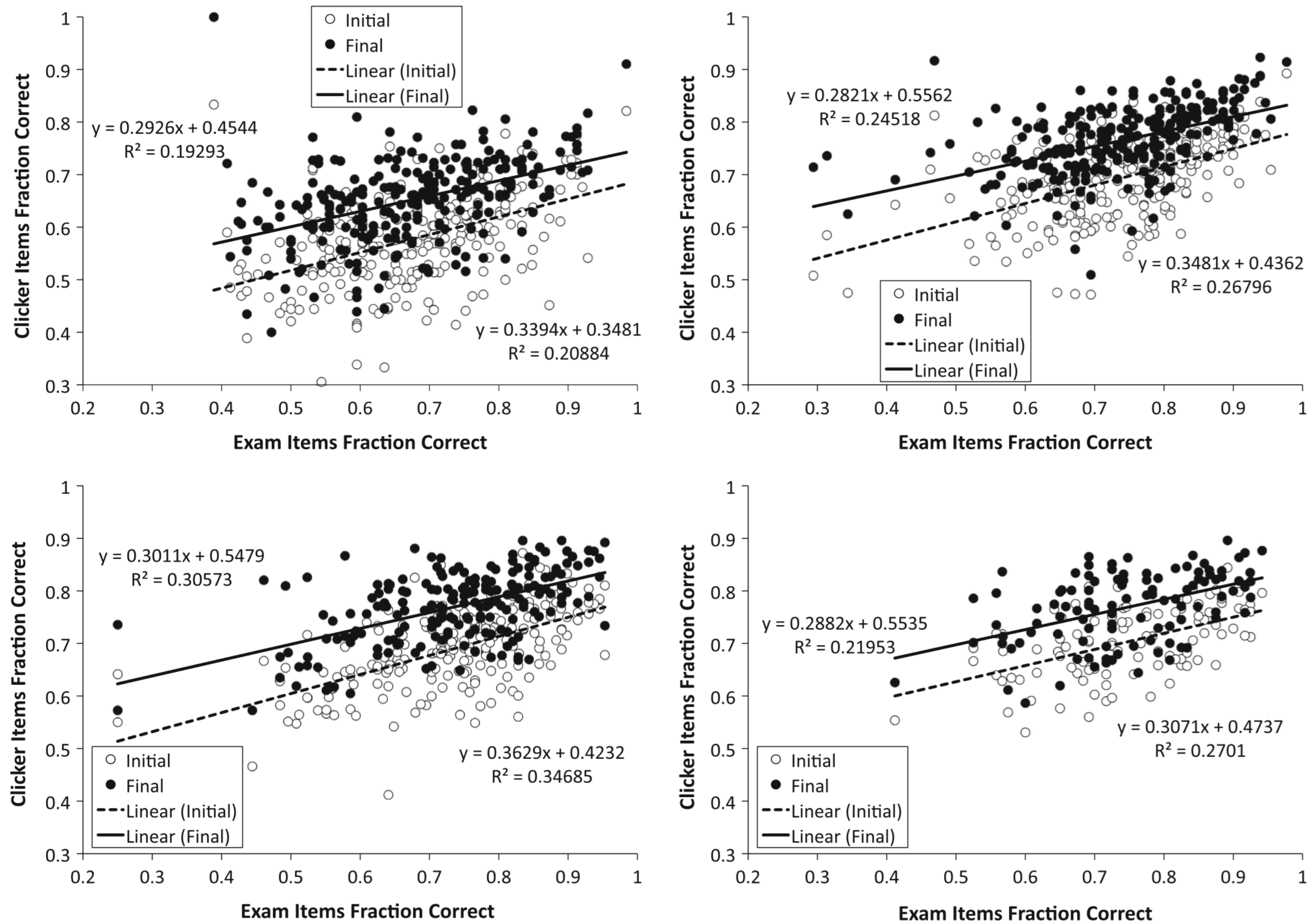
- Graphical input
- Open-ended
- Infinitely many correct answers

As promised: classroom data

- Now some data generated inside the classroom
- Some classical statistics
- Again use IRT to see:
  - How much "random" noise is there?
  - Can problem quality be determined?

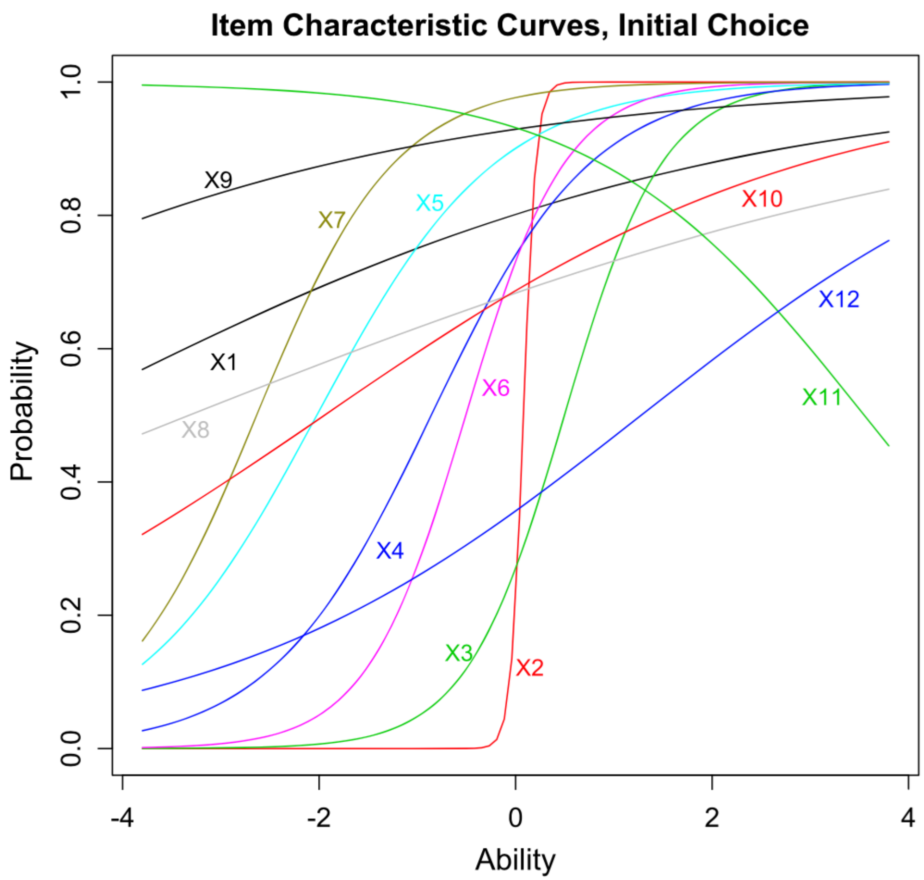
Clicker Data and Exams

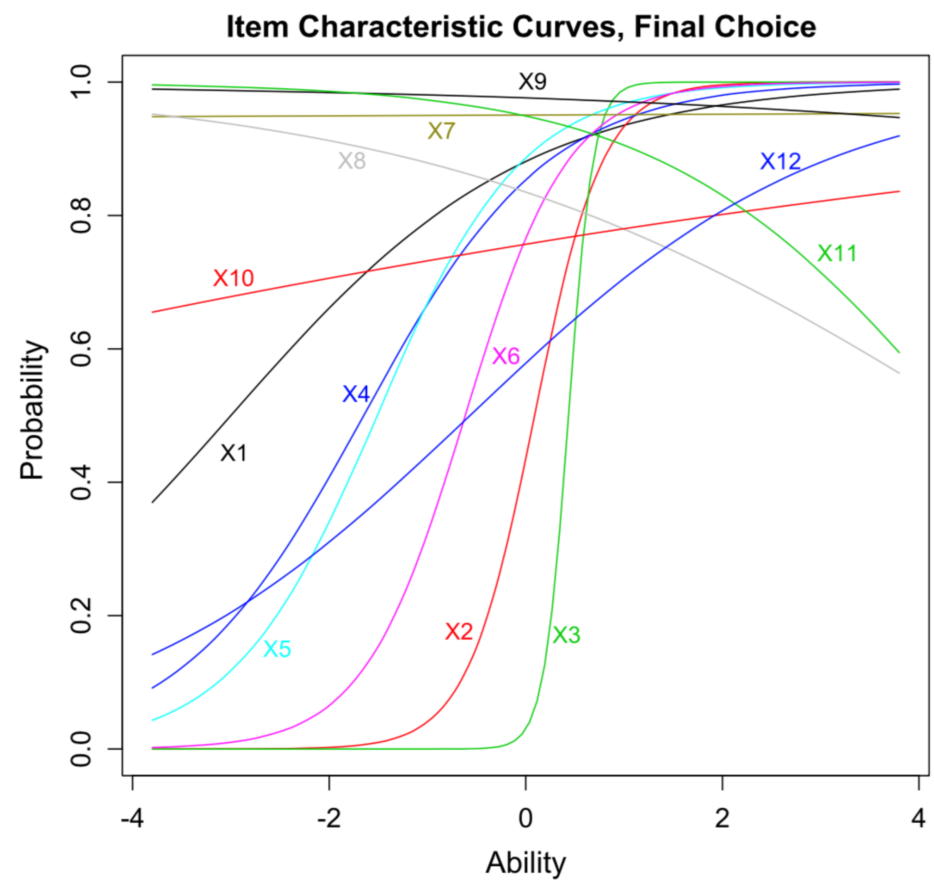
- Is clicker data correlated with exam performance?
  - Initial and final responses equally correlated



Clicker Data IRT

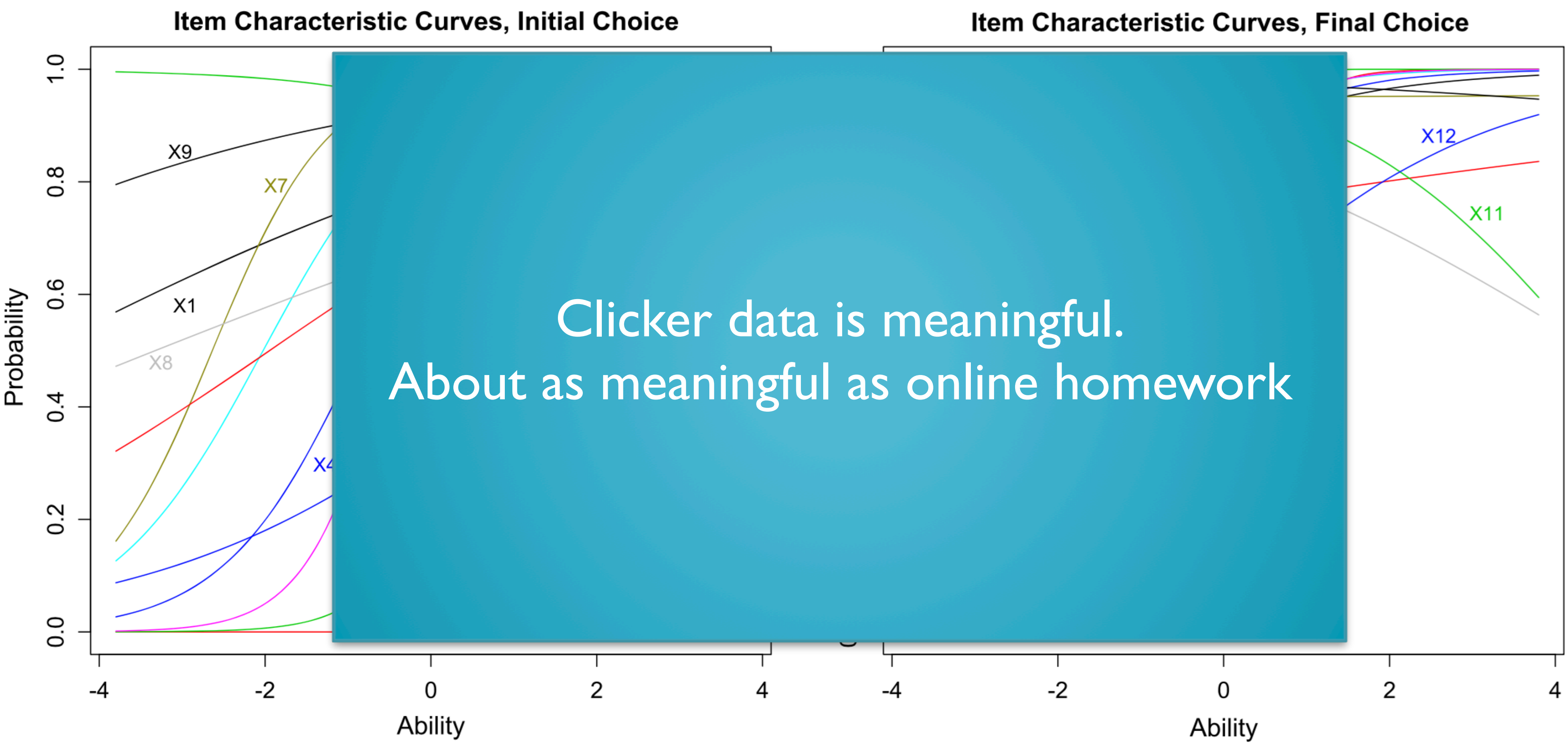
- One lecture (momentum conservation)
  - Initial and final response





Clicker Data IRT

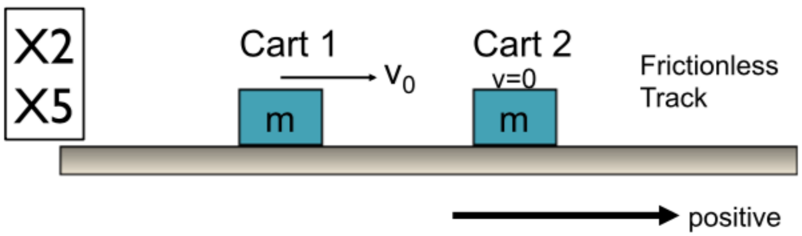
- One lecture (momentum conservation)
  - Initial and final response



Clicker data is meaningful.
About as meaningful as online homework

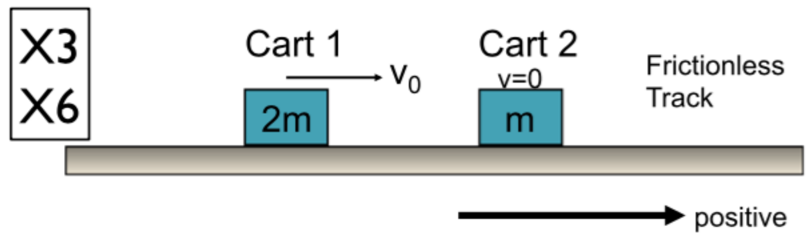
Clicker Data IRT

- "Good" items: much discrimination



Which cart exerts a stronger magnitude force during the collision?

a) Cart 1
b) Cart 2
c) No magnitude forces, both zero
d) Same magnitude forces

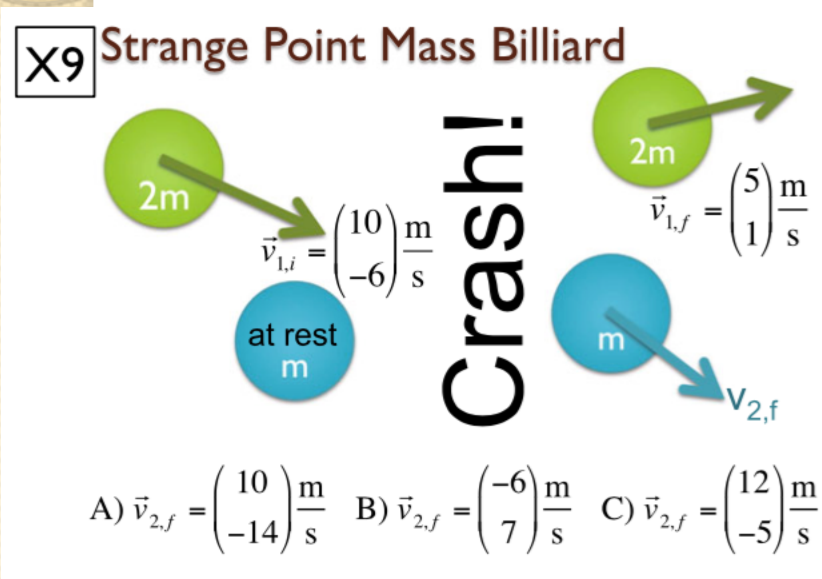


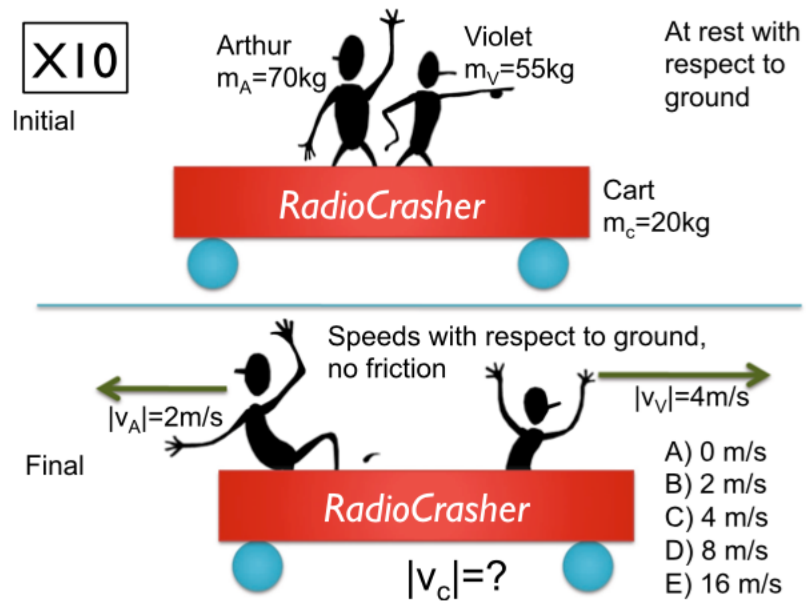
Which cart exerts a stronger magnitude force during the collision?

a) Cart 1
b) Cart 2
c) No magnitude forces, both zero
d) Same magnitude forces

Clicker IRT

- "Bad" problems

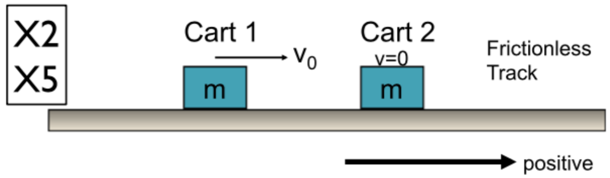




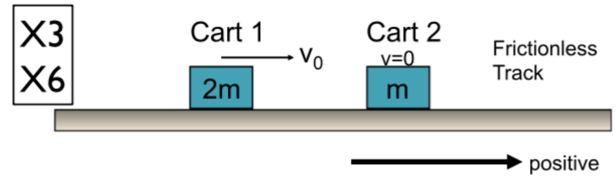
Clicker IRT

- So: what's the difference?

Good

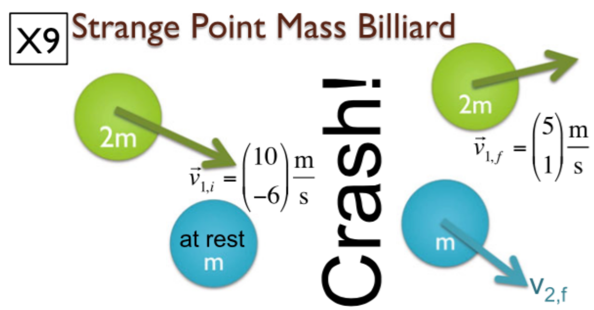


Which cart exerts a stronger magnitude force during the collision?
a) Cart 1
b) Cart 2
c) No magnitude forces, both zero
d) Same magnitude forces

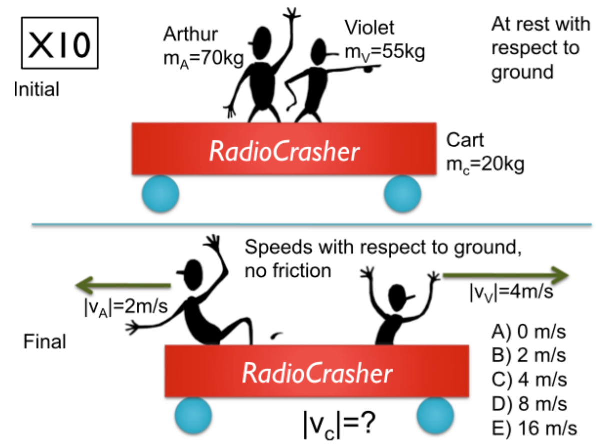


Which cart exerts a stronger magnitude force during the collision?
a) Cart 1
b) Cart 2
c) No magnitude forces, both zero
d) Same magnitude forces

Bad

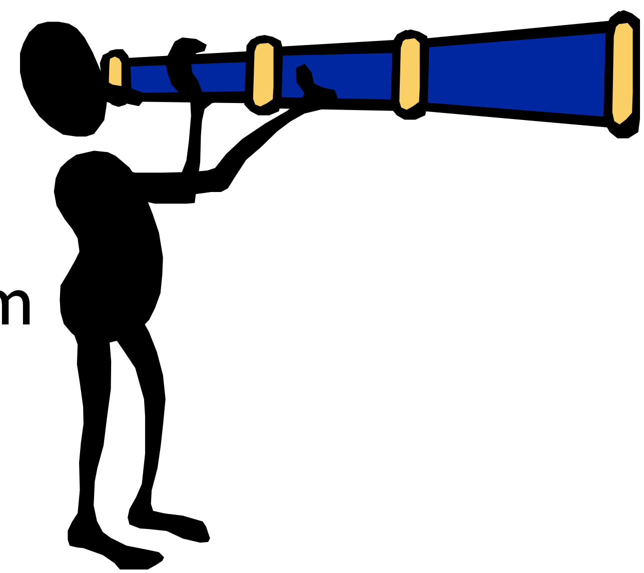


A) $\vec{v}_{2,f} = \begin{pmatrix}10\\-14\end{pmatrix}\frac{\text{m}}{\text{s}}$ B) $\vec{v}_{2,f} = \begin{pmatrix}-6\\7\end{pmatrix}\frac{\text{m}}{\text{s}}$ C) $\vec{v}_{2,f} = \begin{pmatrix}12\\-5\end{pmatrix}\frac{\text{m}}{\text{s}}$



Outlook

- More research needed how problem characteristics influence unproductive behavior
- Looking at the events (and there are millions of them)
  1. Fail on a problem
  2. Do *something*
  3. Succeed on that problem
  - Look at the *something*



Thank you!

- Gerd Kortemeyer
  kortemey@msu.edu