Multiple Tries on Trial

Gerd Kortemeyer

Michigan State University

AAPT Conference
July 2015

The Accusation

- Online homework fosters unproductive behavior
- Too many multiple tries are at fault

… or maybe …

Too few tries are at fault

Typical Online Physics Problem

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 » ... omentum and Collisions ◄ Timer Notes Stored Links Evaluate Feedback Print Info

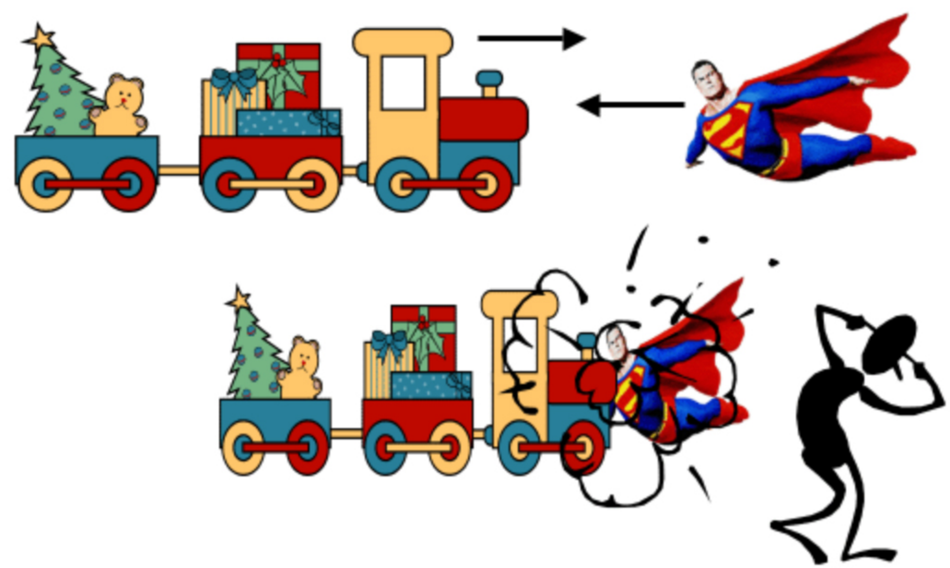
Functions Edit ... Grades Content Settings

Randomized problem

Superman Stops ... n

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

An out-of-control train is racing tow... is 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?



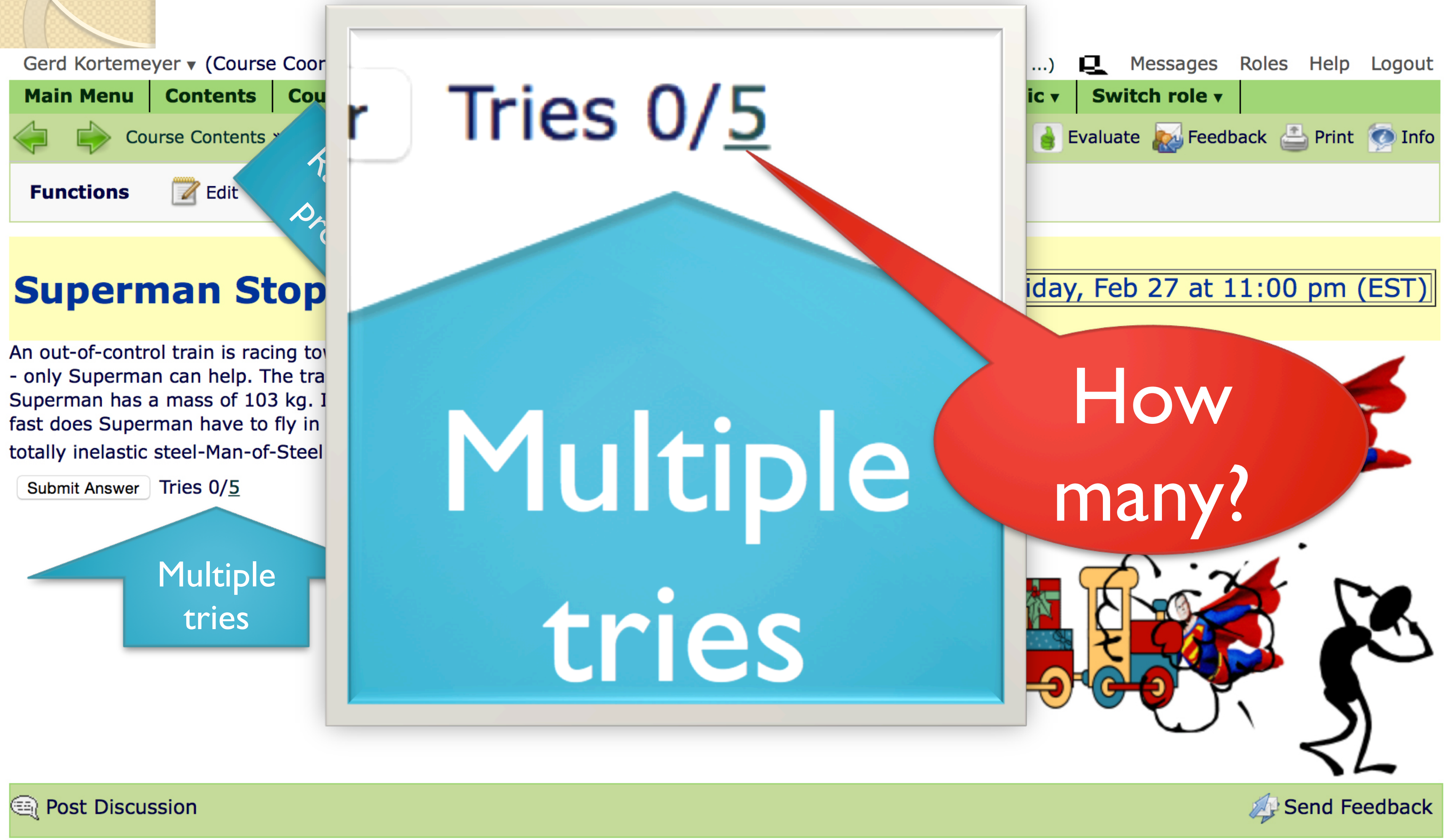
Submit Answer Tries 0/5

Multiple tries

Open-ended numerical

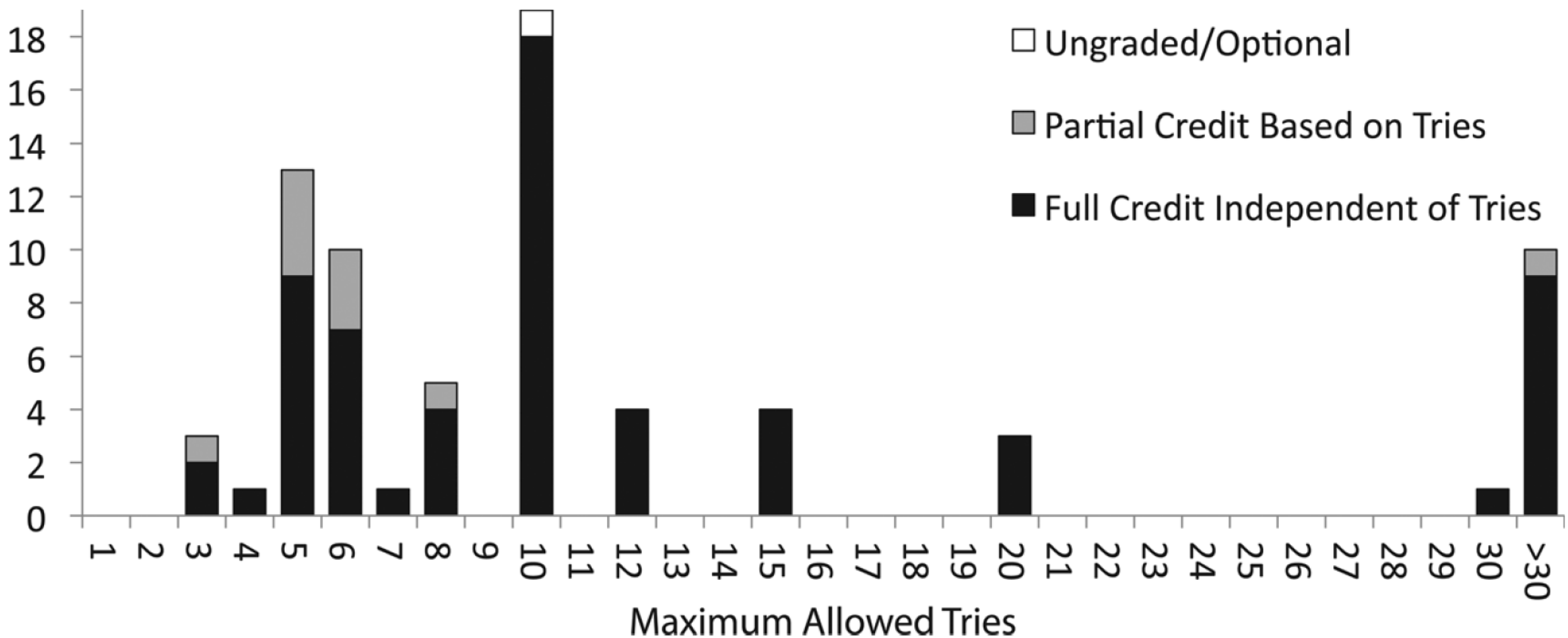
Post Discussion Send Feedback

Typical Online Physics Problem



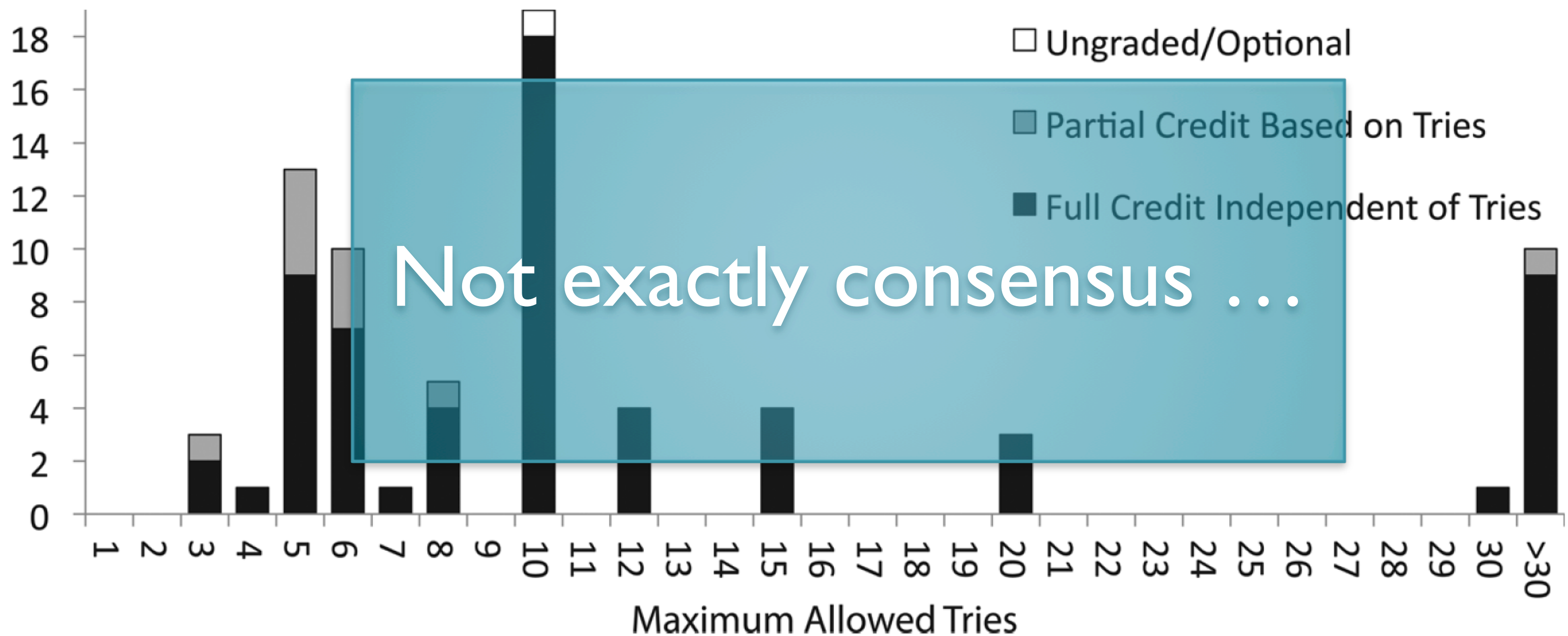
How Many Tries to Grant?

- 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

Unproductive Behavior

- Random Guessing
  - Submitting "random" guesses to online homework
  - Possibly more likely if more tries are allowed
    - Not taking attempts seriously
- Copying or Very Closely Collaborating
  - Submitting other people's work to online homework
  - Possibly more likely if less tries are allowed
    - "only chance to get the points"

Unproductive Behavior

- Random Guessing
  - Submitting "random" guesses to online homework
  - Possibly more likely if more tries are allowed
    - Not taking attempts seriously
- Copying or Very Closely Collaborating
  - Submitting other people's work to online homework
  - Possibly more likely if less tries are allowed
    - "only chance"

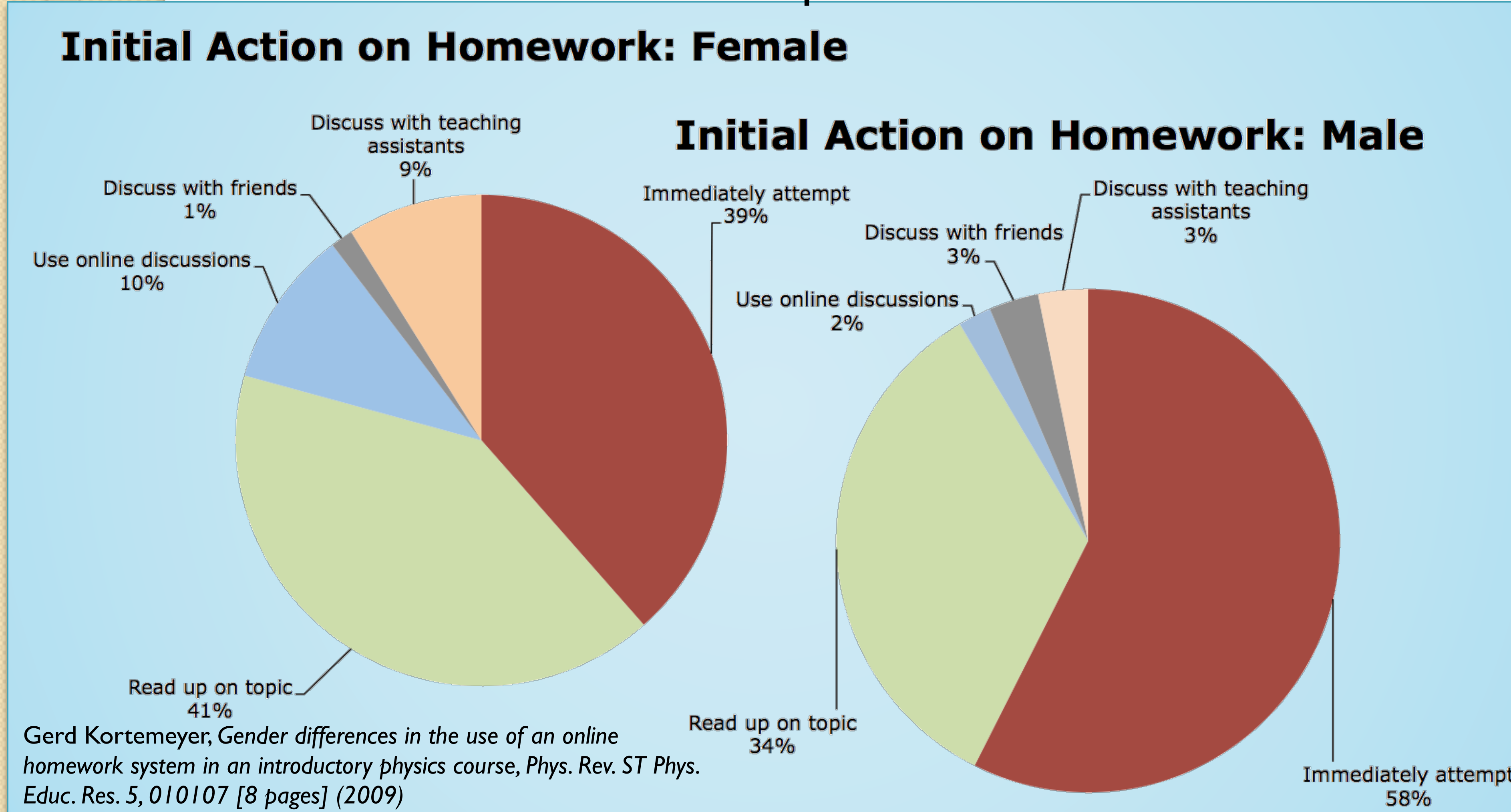
How do you really know?

Unproductive Behavior

- How do you really know what’s happening?
- Ask the students
  - Surveys
- “Ask” the online homework systems
  - Logging every transaction
    - Time stamps
    - Correct/incorrect
    - Allowed number of attempts

Survey

- What do students do when they first encounter a new "unknown" homework problem?



Gerd Kortemeyer, *Gender differences in the use of an online homework system in an introductory physics course, Phys. Rev. ST Phys. Educ. Res. 5, 010107 [8 pages] (2009)*

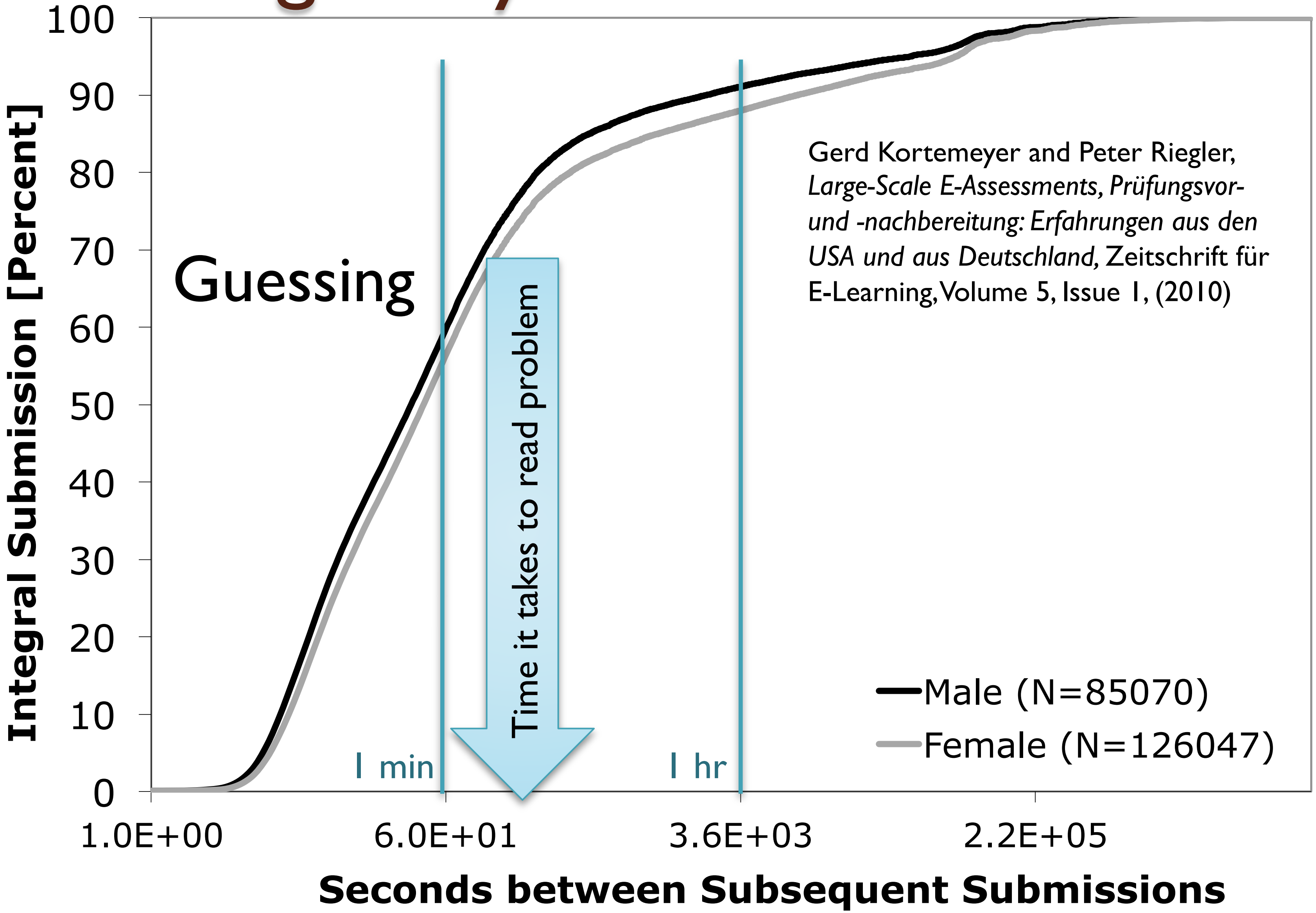
Survey

- Immediately attempt – “guessing?”:
  - Male students: 58%
  - Female students: 39%
- Discuss with friends or online – “copying?”:
  - Male students: 5%
  - Female students: 11%
- Stereotypical: “Real men don’t ask for directions”

Survey versus "Hard" Data

- There is definitely the danger of guessing or collaborating too closely
- But self-reported data is notoriously unreliable
- What is in the data logs?
  - Timing analysis
  - Tries versus success
  - Data mining
  - Item Response Theory

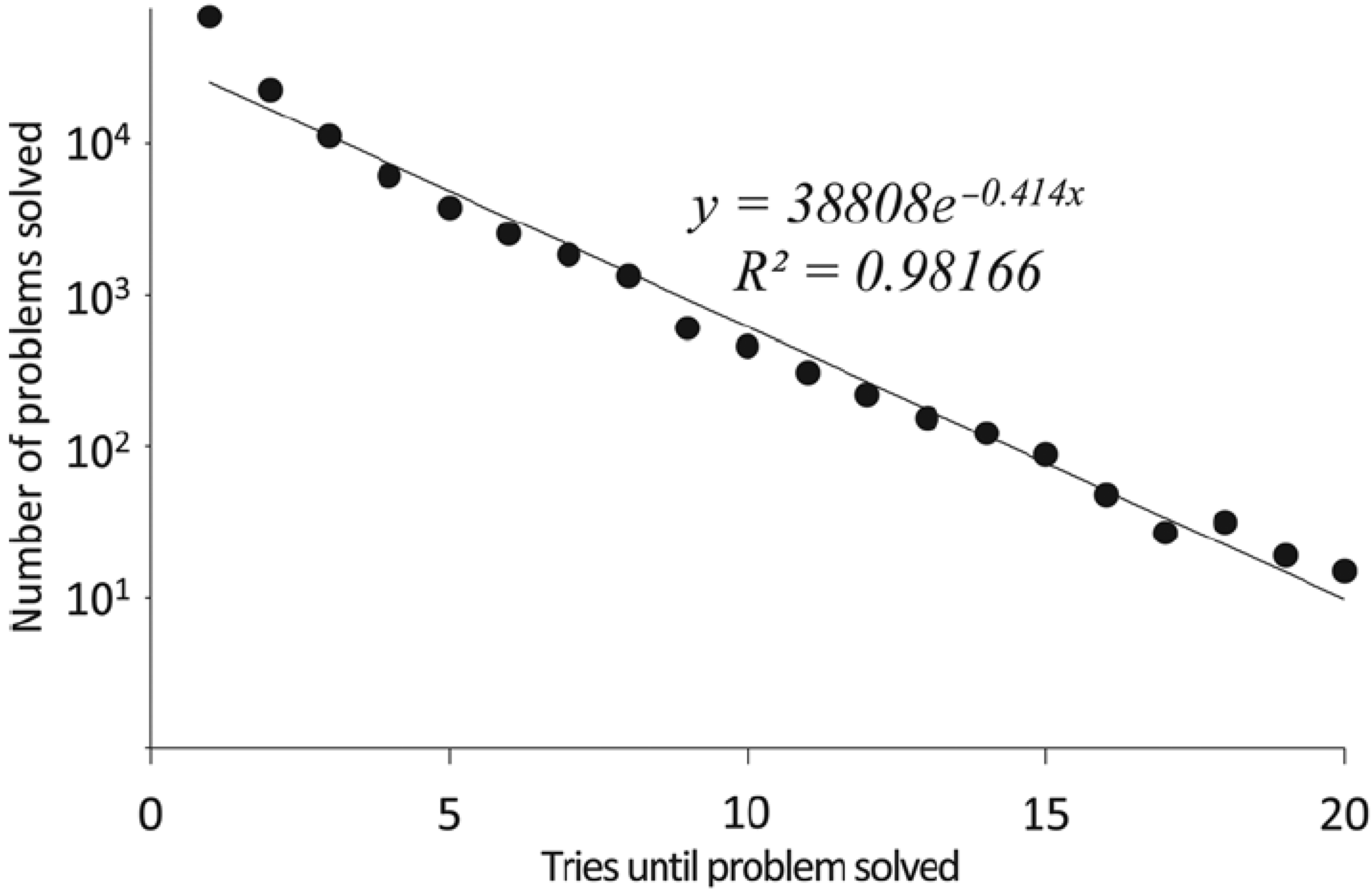
Timing Analysis



Gerd Kortemeyer and Peter Riegler, *Large-Scale E-Assessments, Prüfungsvor- und -nachbereitung: Erfahrungen aus den USA und aus Deutschland,* Zeitschrift für E-Learning, Volume 5, Issue 1, (2010)

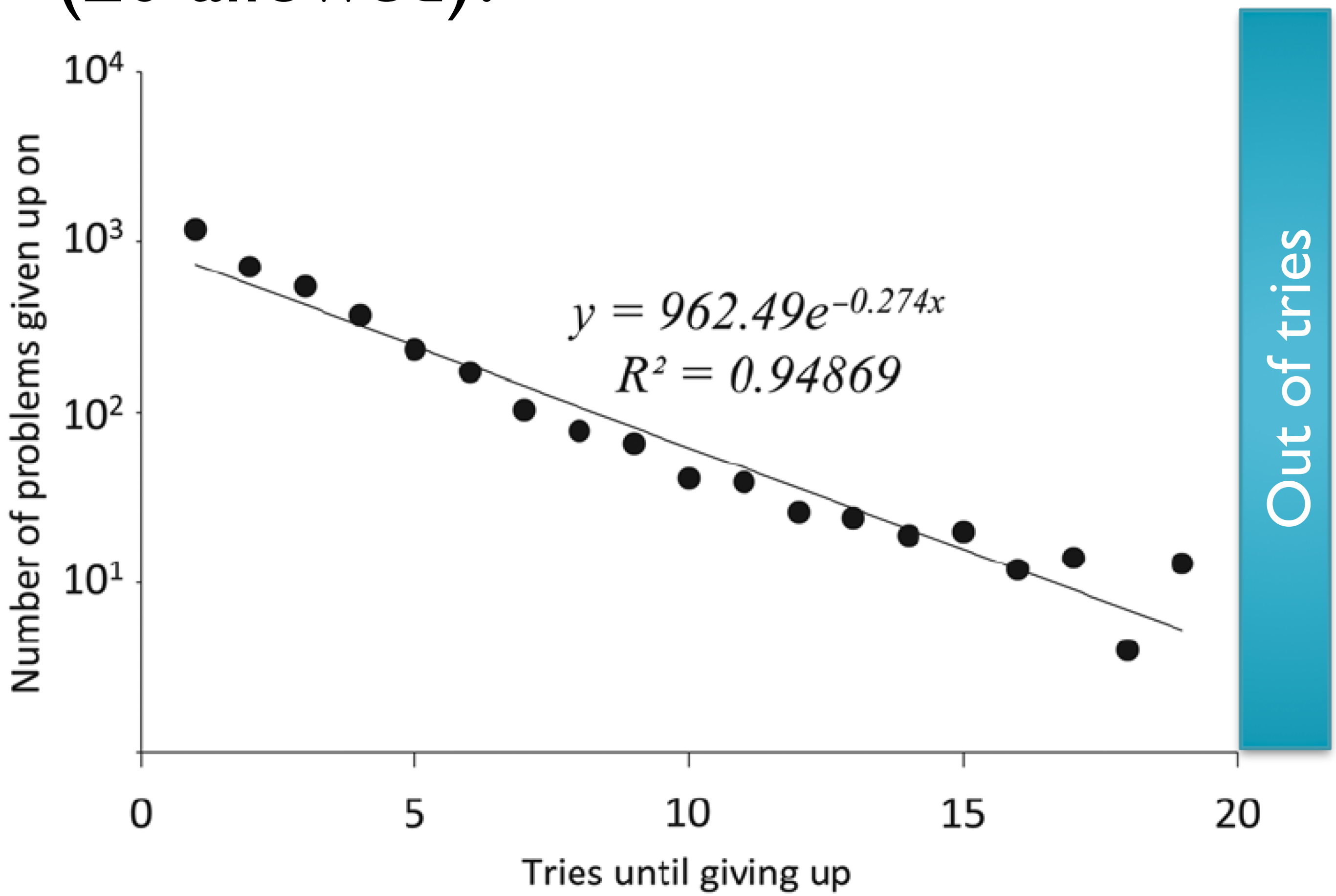
Tries versus Success

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



Tries versus Success

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



Tries versus Success

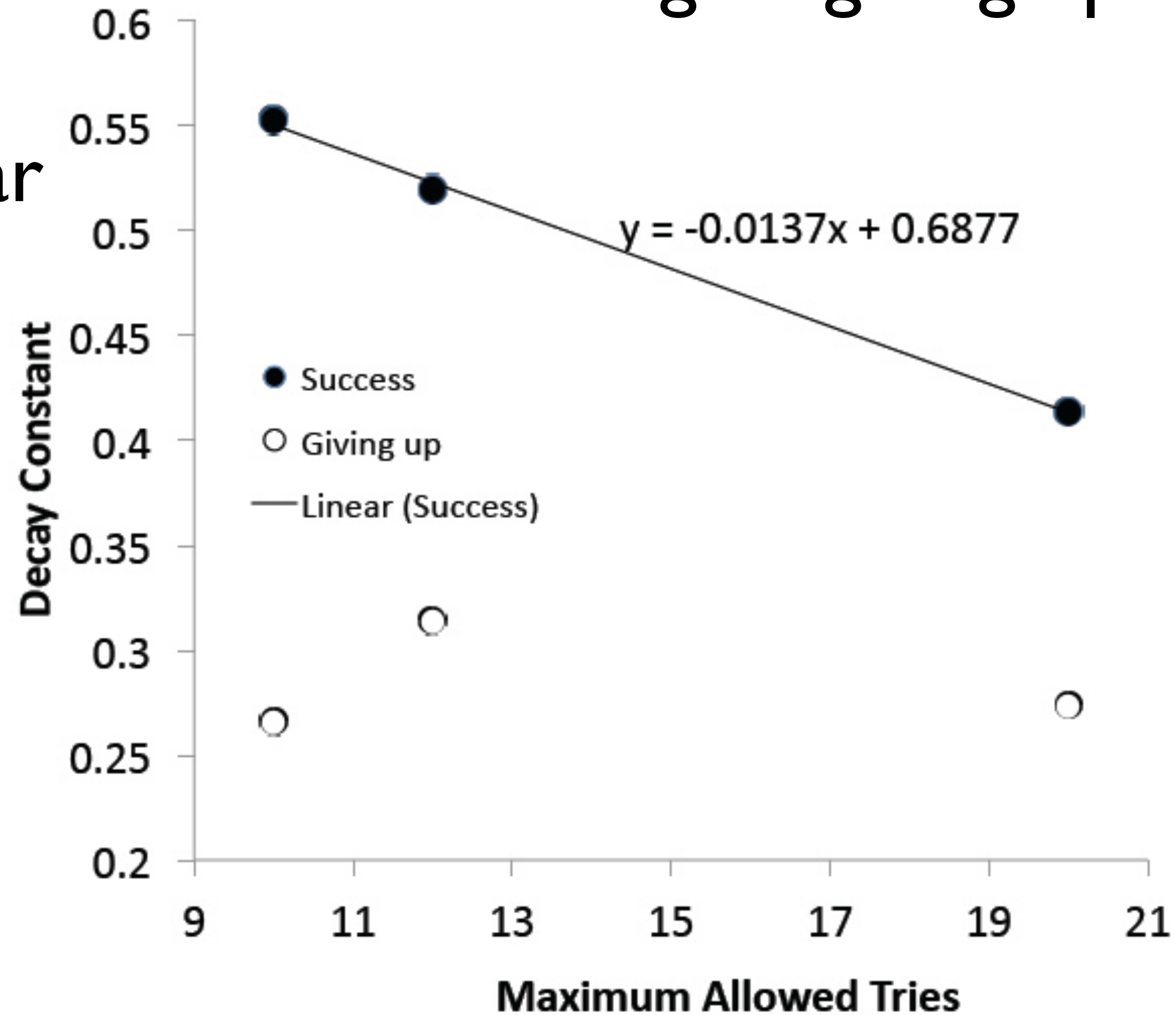
- Comparing three classes:
  10 tries, 12 tries, and 20 tries max.
- Surprisingly, for all classes, both success and giving up follow

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

- Tries are independent of each other!
- Lambdas are like probabilities

Tries versus Success

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

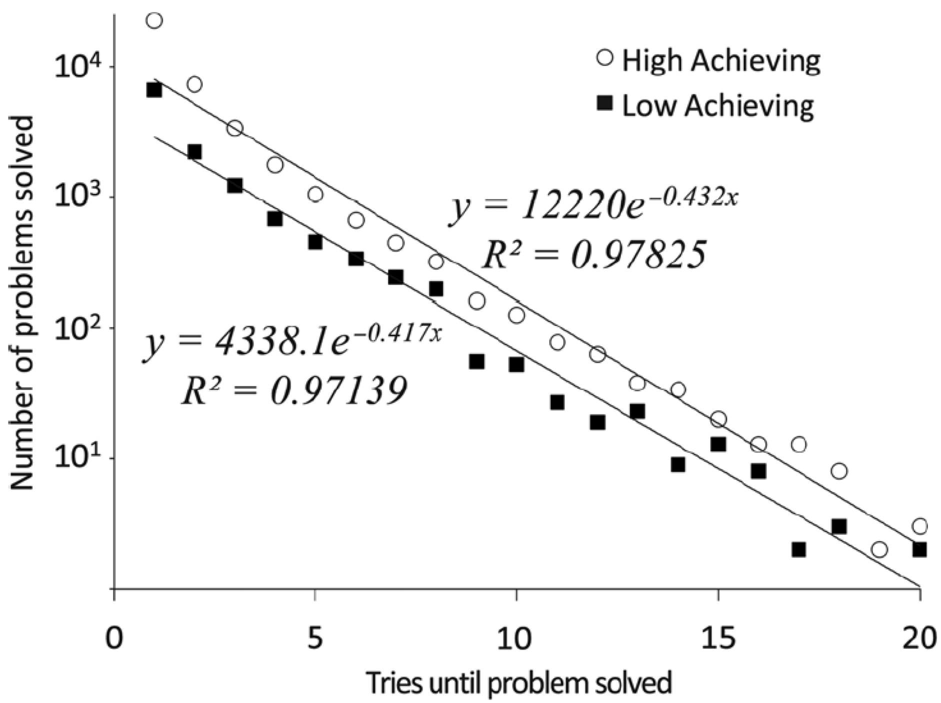


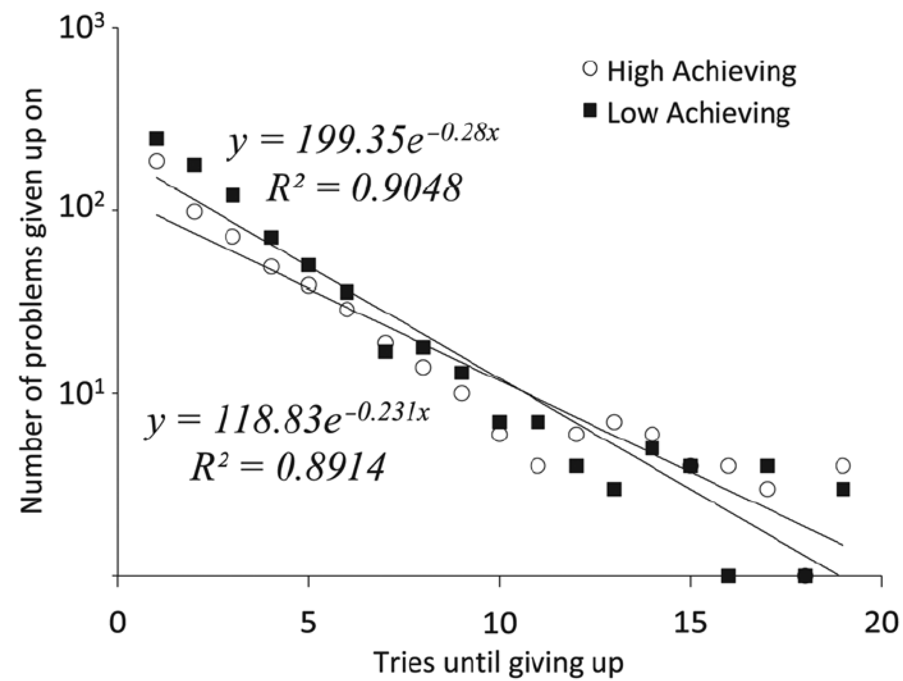
Tries versus Success

- Following "decay" law:
  - students do not really profit from earlier tries
  - students do no learn from their mistakes
- Giving more tries reduces the probability of success on a particular try
- Also: total amount of successfully solved homework remains about the same, independent of number of allowed tries
  - Running out of tries is rare

Tries versus Success

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

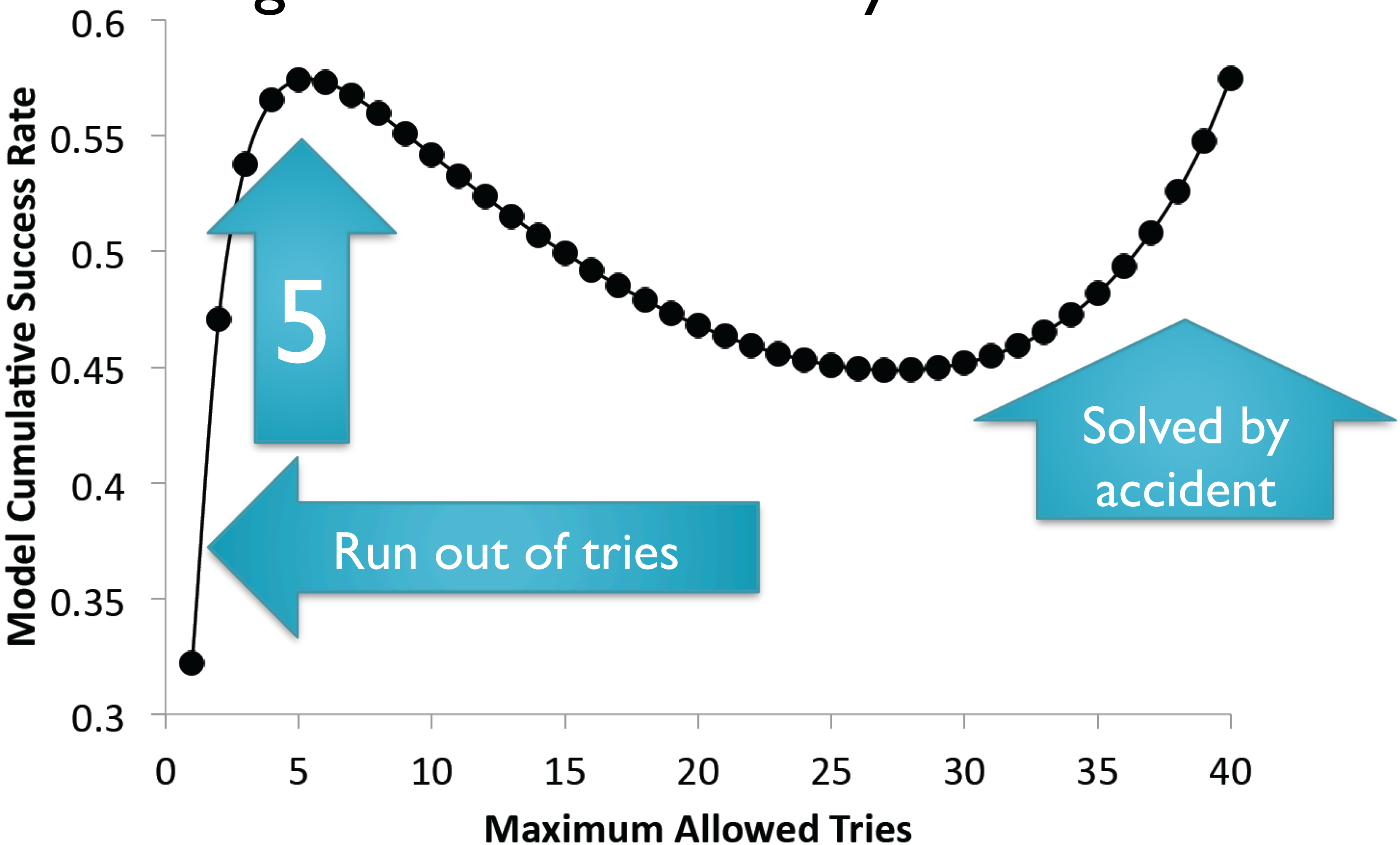




No.

Tries versus Success

- Using this model of "decay constants"



Data Mining Access Logs

- Is guessing and copying important?
- What behavior leads to which grade?
- Define behavioral features
  - Extract from logs
- Define performance classes
- Go!

Data Mining Access Logs

- 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

Data Mining Access Logs

Performance classes, as fine-grained as you want:

TABLE 1

SELECTING 9 CLASS LABELS REGARDING TO STUDENTS' GRADES

Class	Grade	Student #	Percentage
1	0.0	2	0.9%
2	0.5	0	0.0%
3	1.0	10	4.4%
4	1.5	28	12.4%
5	2.0	23	10.1%
6	2.5	43	18.9%
7	3.0	52	22.9%
8	3.5	41	18.0%
9	4.0	28	12.4%

TABLE 2

SELECTING 3 CLASS LABELS REGARDING TO STUDENTS' GRADES

Class	Grade	Student #	Percentage
High	Grade >= 3.5	69	30.40%
Middle	2.0 < Grade < 3.5	95	41.80%
Low	Grade <= 2.0	63	27.80%

TABLE 3

SELECTING 2 CLASS LABELS REGARDING TO STUDENTS' GRADES

Class	Grade	Student #	Percentage
Passed	Grade > 2.0	164	72.2%
Failed	Grade <= 2.0	63	27.80%

Data Mining Access Logs

- See how much you can explain

		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**

Data Mining Access Logs

- … and find the 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

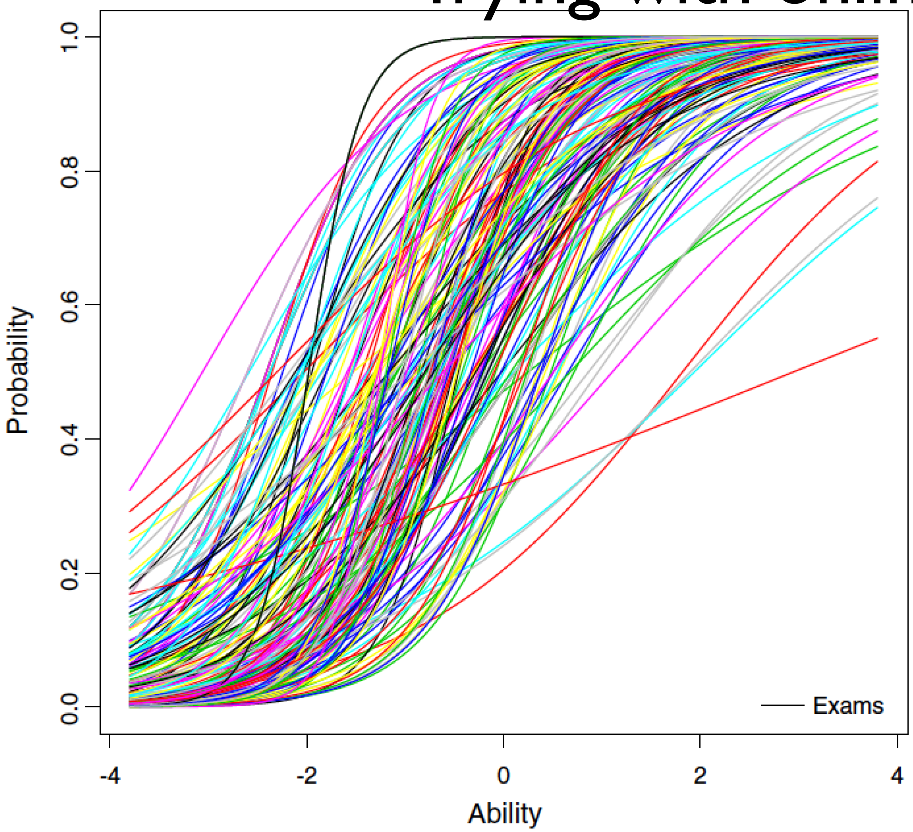
Data Mining Access Logs

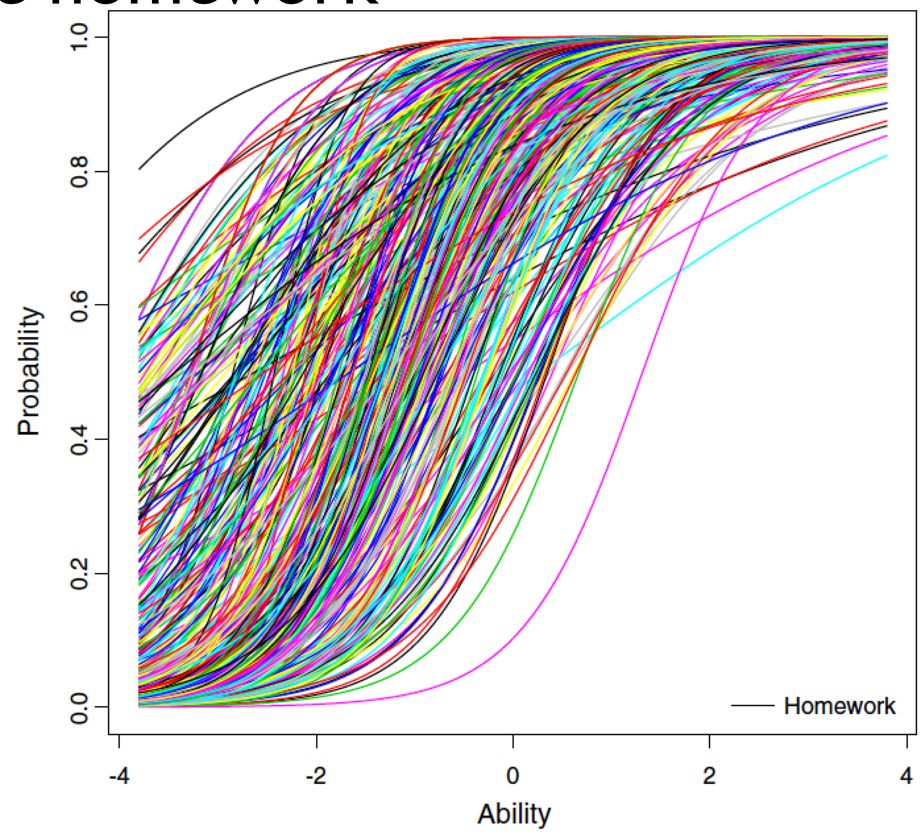
- What does that mean?
  - Most important: did the student solve homework problems eventually?
  - Second: not too many tries
  - Third (factor four lower!): did they get it right on the first attempt?
- Tenacity more important than immediate genius!

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

Item Response Theory

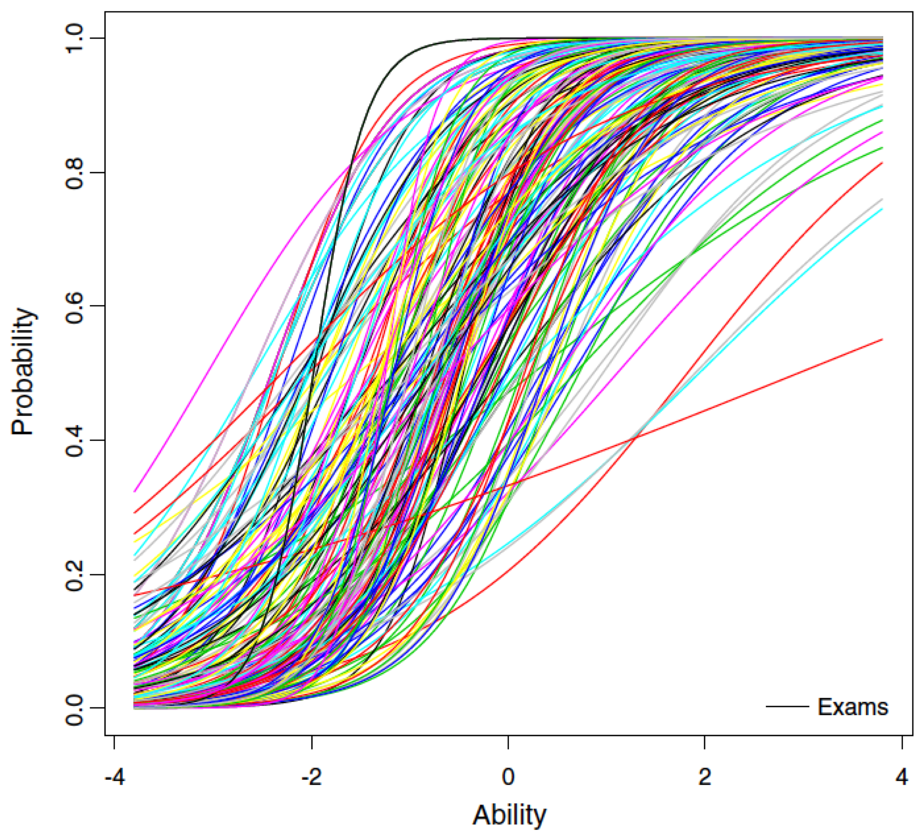
- IRT was developed for summative assessments
  - Trying with online homework

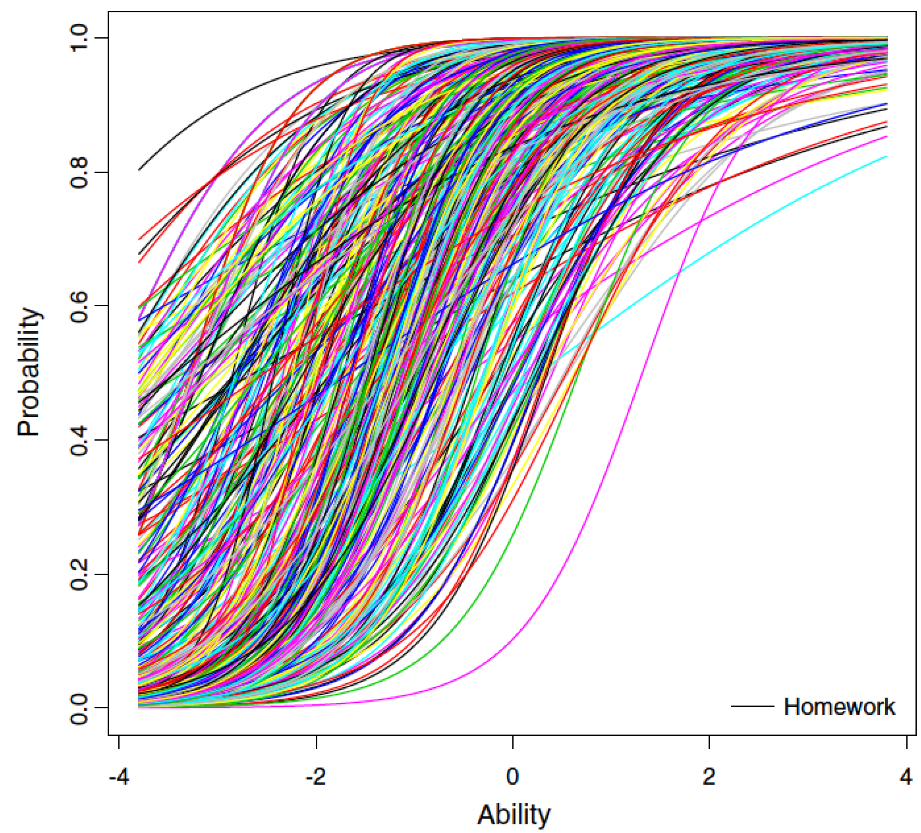




Item Response Theory

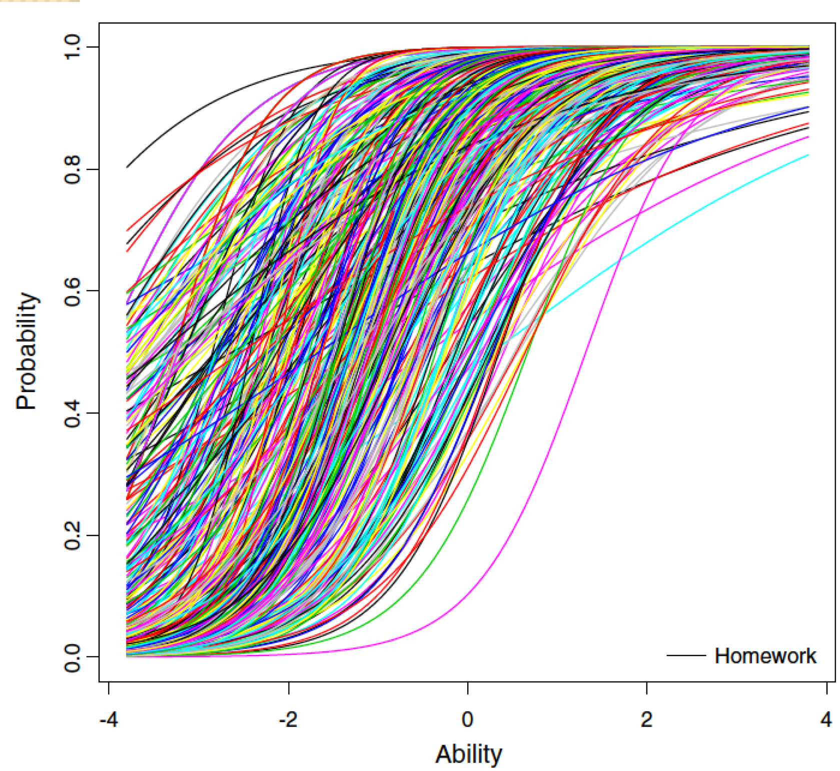
- You can see the "noise"
- This is guessing and copying

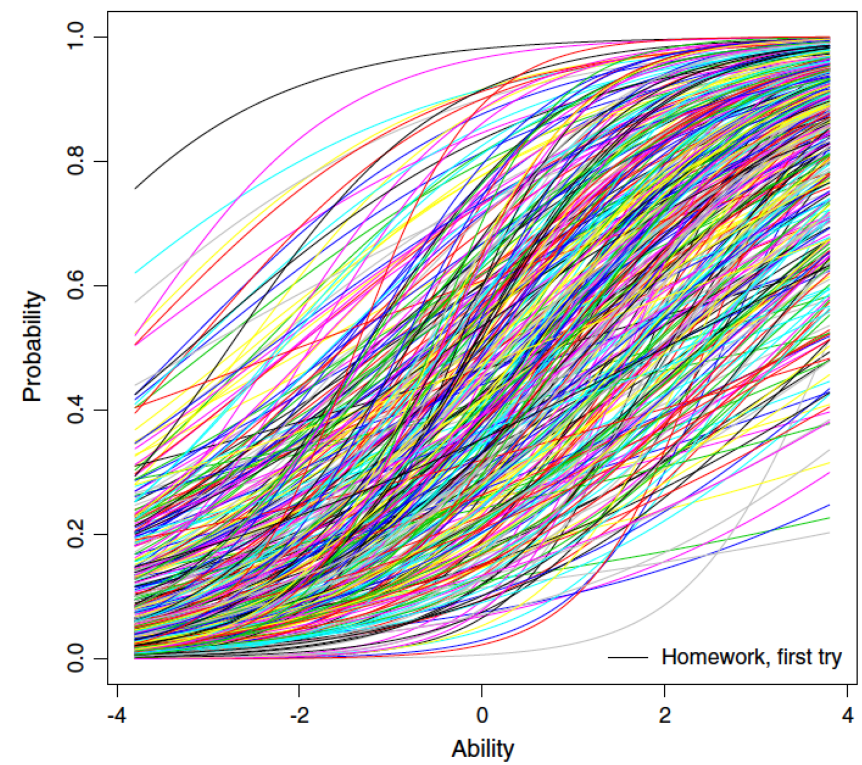




Item Response Theory

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





Item Response Theory

- 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)

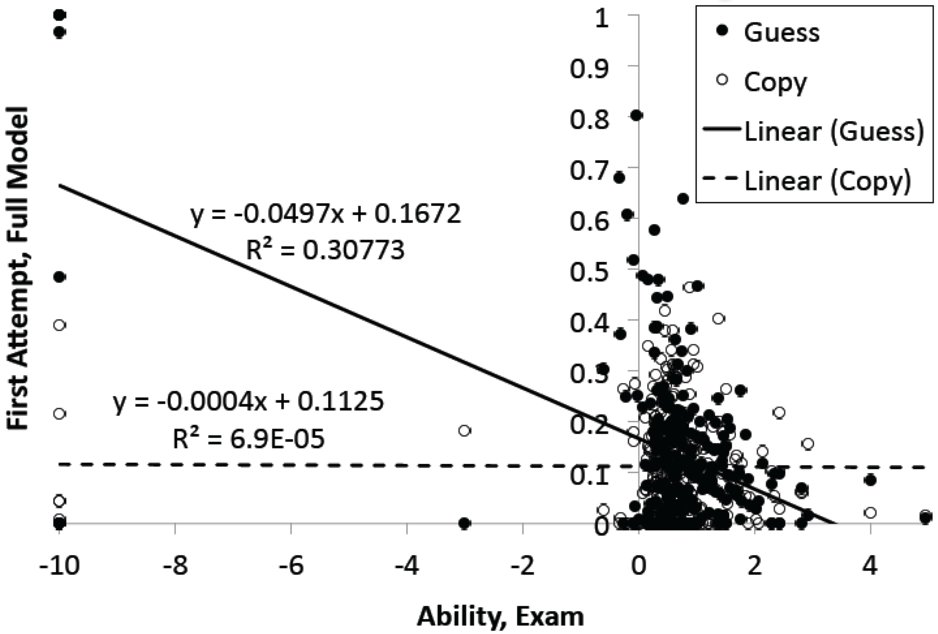
Item Response Theory

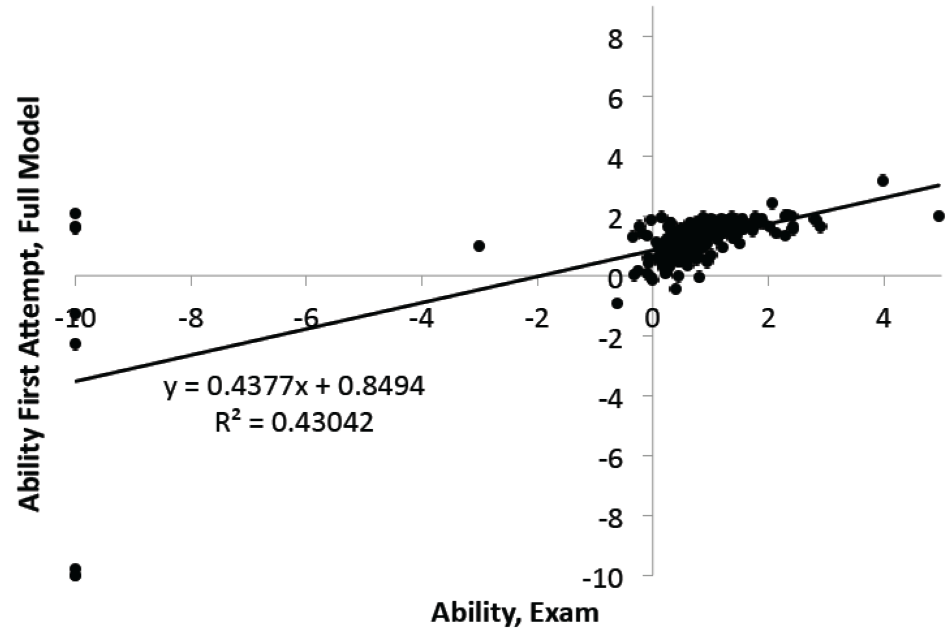
- Modeling unproductive behavior
- Need new IRT model

$$
\begin{aligned}
\tilde{p}_{ij} &= \chi_j (1 - p_{ij}) + (1 - \gamma_j) p_{ij} \\
&= \chi_j + \frac{1 - \gamma_j - \chi_j}{1 + \exp(a_i(b_i - \theta_j))}
\end{aligned}
$$

- Guessing and copying as learner traits

Item Response Theory





Emre Gönülateş and Gerd Kortemeyer, *A New Item Response Theory Model for Open-Ended Online Homework with Multiple Allowed Attempts*, Phys. Rev. ST Phys. Educ. Res. (submitted)

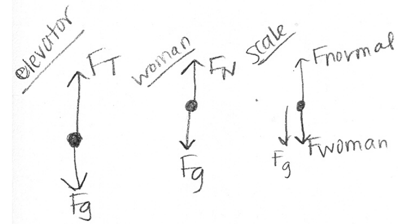
- Taking unproductive behavior into account increases predictive power
- Students of all exam abilities copy
- Better students guess less
- Copying strong component of first-try success

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) $F = ma$

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

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

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

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

$F_N - 60(9.8) = 57$

$F_N = 645\ N$

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

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

$a = 8.79\ m/s^2$

$F_N - F_g = 60(8.79)$

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

$F_N - 117.6 = 527.4$

$\boxed{F_N = 645\ N}$

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

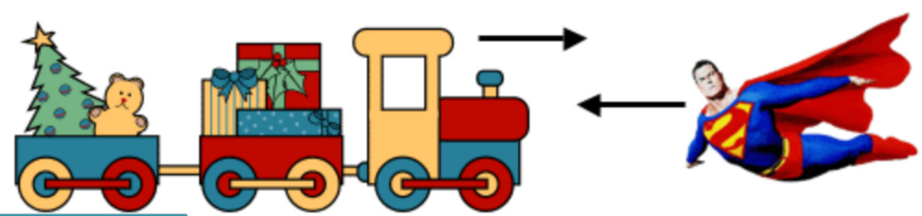
Functions Edit Content Grades Content Settings

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?

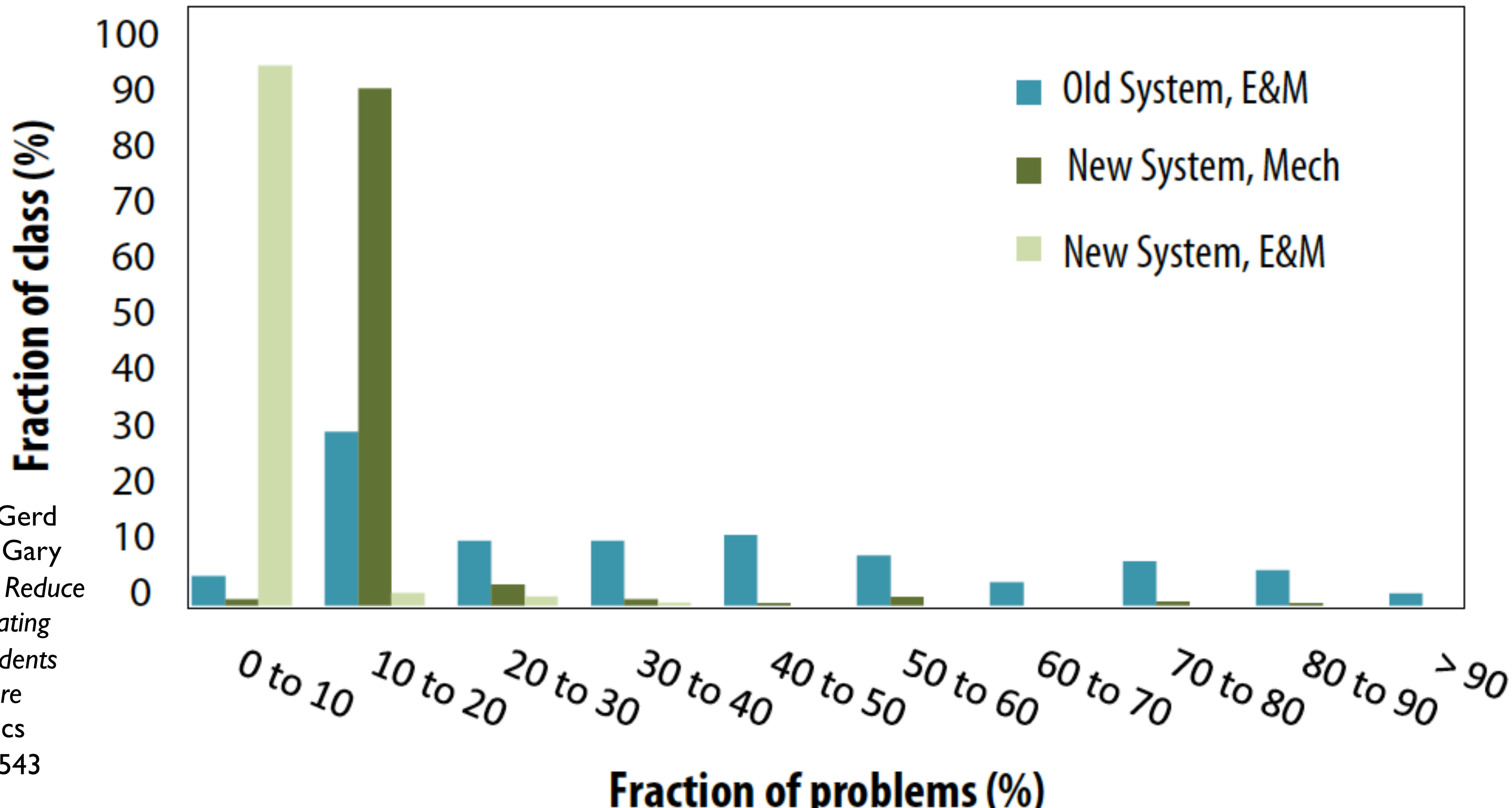
Send Feedback

So?

- We saw: copying and guessing are clearly present
- Is there anything that can be done?
- Idea: make formative assessment more effective by increasing the number of summative assessment venues
  - More exams
  - Intro physics course, before/after

More Exams

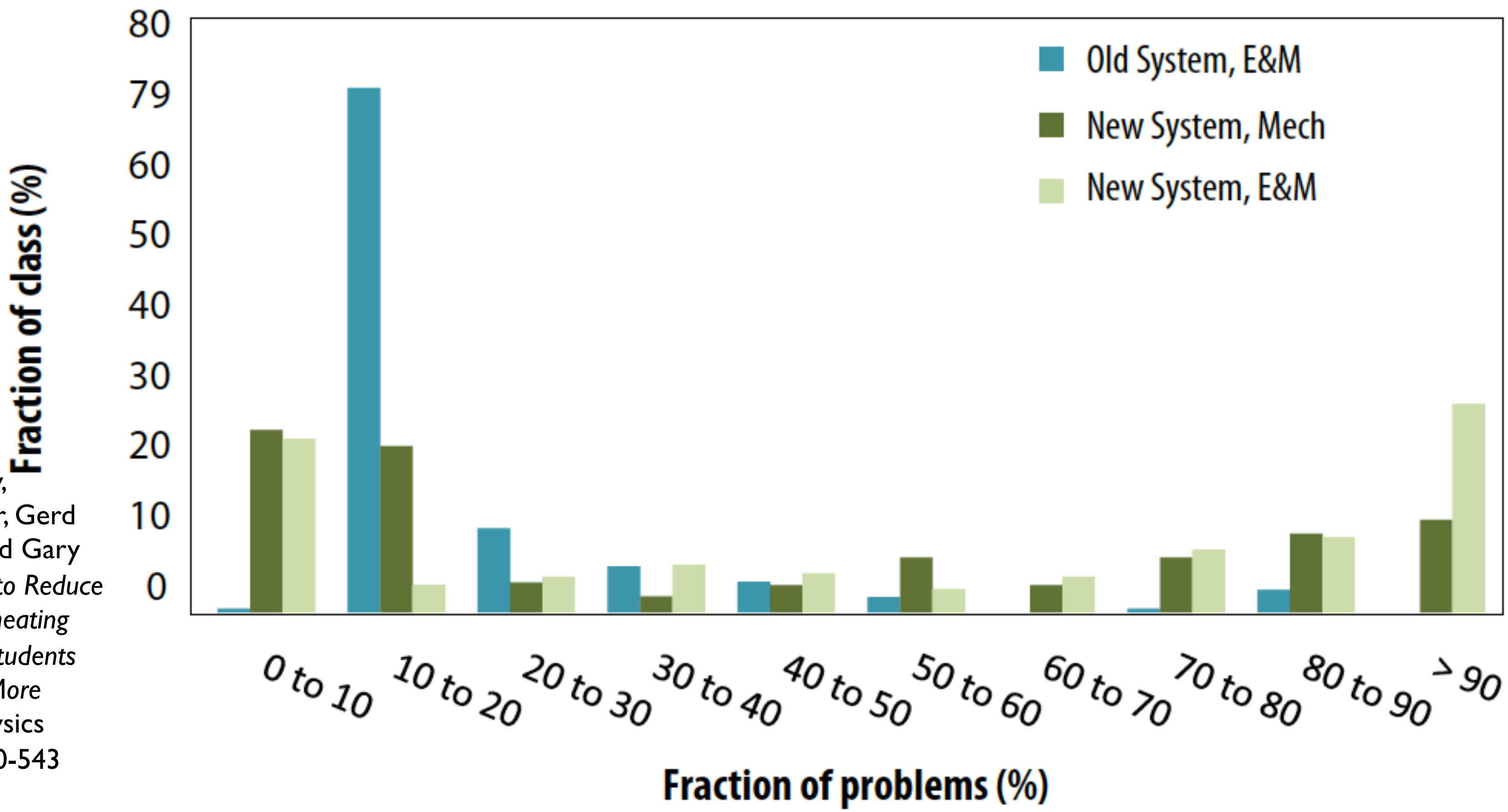
- Self-reported use of 3rd party cheat sites, which students use to copy answers



James T. Laverty, Wolfgang Bauer, Gerd Kortemeyer, and Gary Westfall, *Want to Reduce Guessing and Cheating While Making Students Happier? Give More Exams!*, The Physics Teacher **50**, 540-543 (2012)

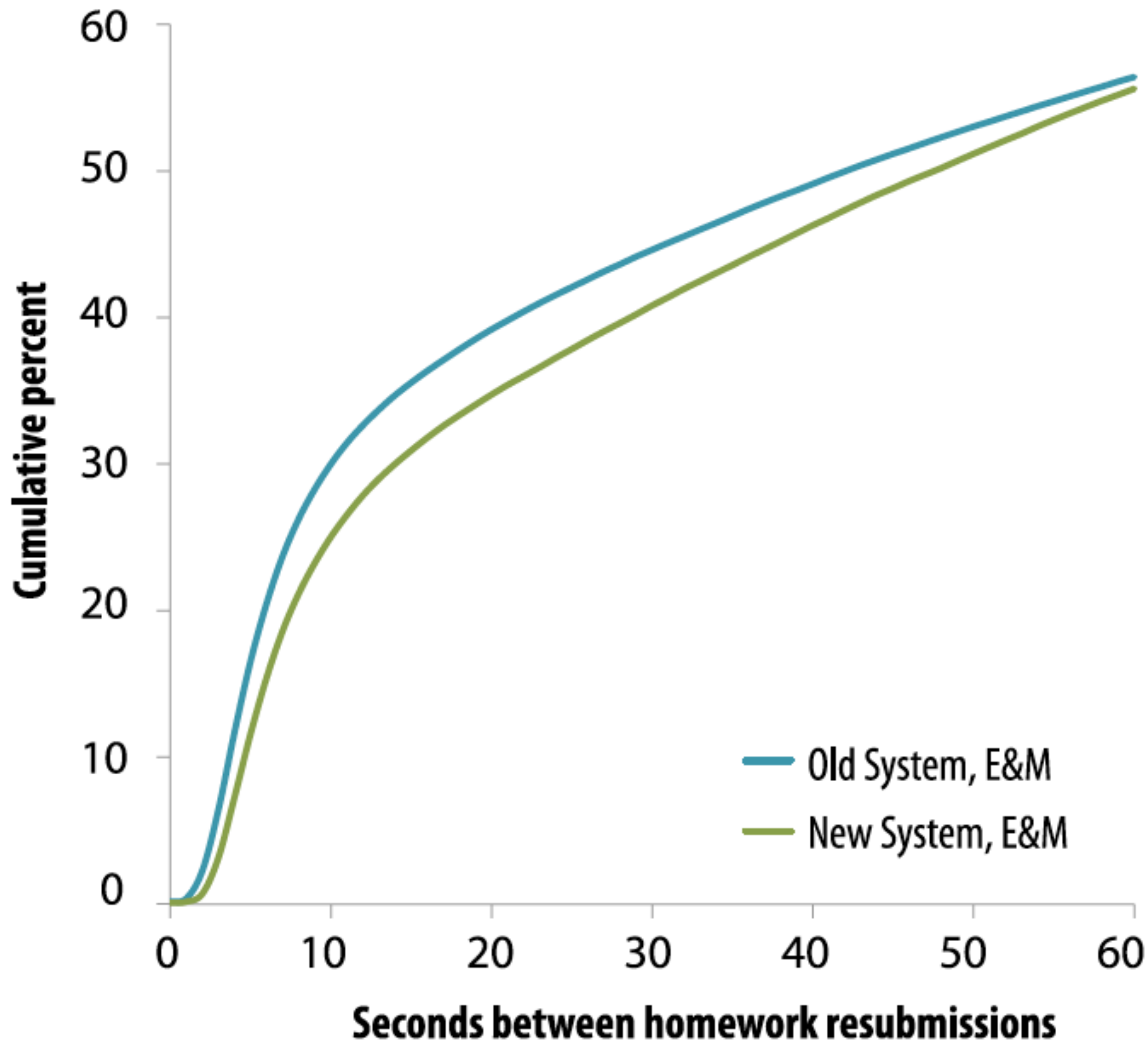
More Exams

- Sanctioned internal discussions, where course instructors participate



James T. Laverty, Wolfgang Bauer, Gerd Kortemeyer, and Gary Westfall, *Want to Reduce Guessing and Cheating While Making Students Happier? Give More Exams!*, The Physics Teacher **50**, 540-543 (2012)

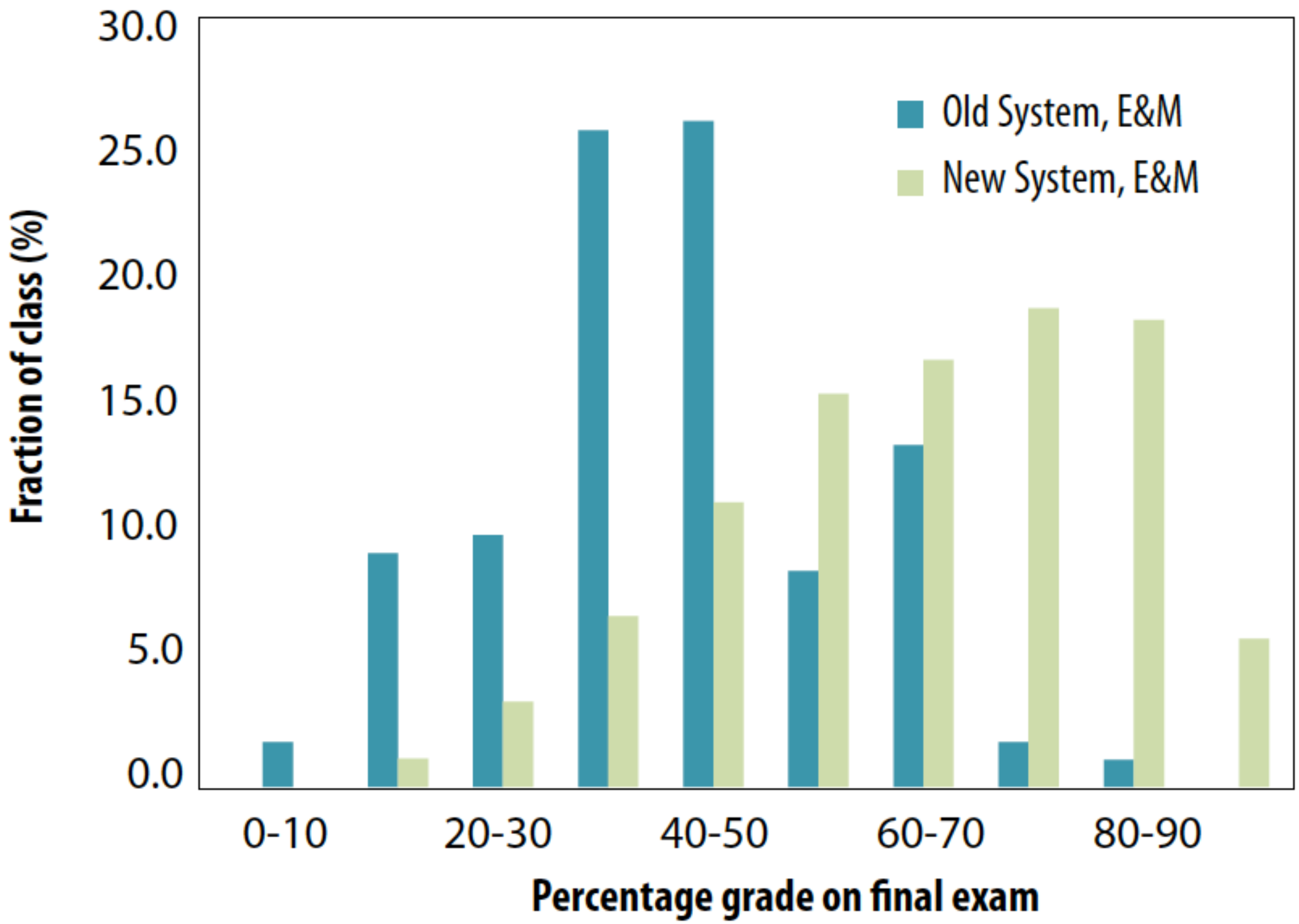
More Exams



James T. Laverty, Wolfgang Bauer, Gerd Kortemeyer, and Gary Westfall, *Want to Reduce Guessing and Cheating While Making Students Happier? Give More Exams!*, The Physics Teacher **50**, 540-543 (2012)

More Exams

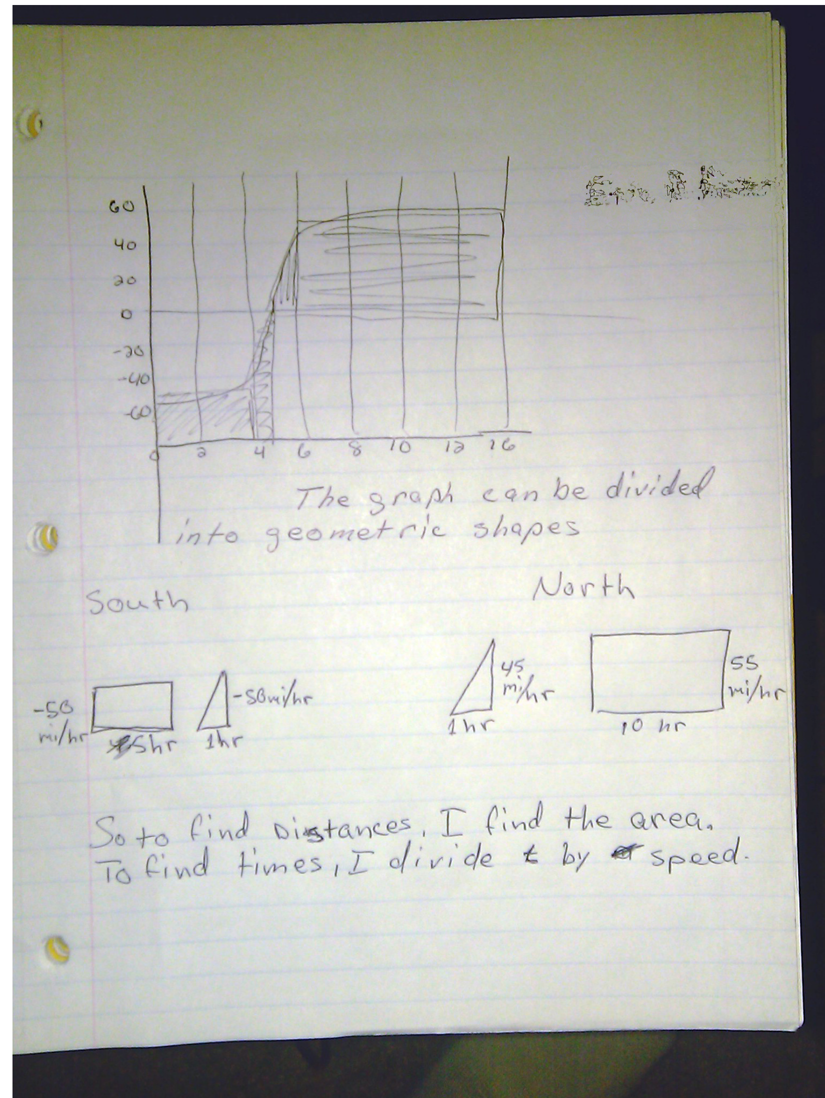
- The proof is in the pudding: Final Exam



James T. Laverty, Wolfgang Bauer, Gerd Kortemeyer, and Gary Westfall, *Want to Reduce Guessing and Cheating While Making Students Happier? Give More Exams!*, The Physics Teacher **50**, 540-543 (2012)

Another Approach

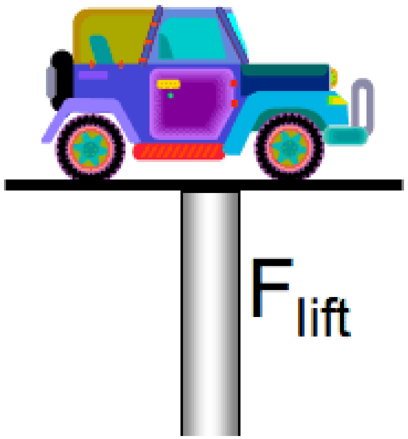
- Curb plug-and-chug
- Have students turn in some derivations and graphs simply by photographing them with their cell phones and uploading them to the CMS
    - Maybe we don't know how to do that, but they sure do!



… 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 …



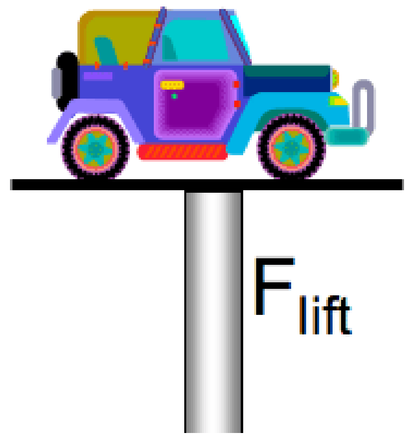
A car (mass of 750

sitting on a car lift in a shop (neglect th

of the lift itself). While the car is being l

up, it is speeding up with 2.3 m/s^2. Wh

the magnitude of the lifting force?



A car (mass of 990 kg) is sitting on a car lift in a shop (neglect the mass of the lift itself). While the car is being lowered, it is speeding up with 3.3 m/s^2. What is the magnitude of the lifting force?

A car (mass of 940 kg) is sitting on a car lift in a shop (neglect the mass of the lift itself). While the car is being lifted up, it is slowing down with 2.1 m/s^2. What is the magnitude of the lifting force?

Lifting/lowering,
speeding up/slowing down,
different numbers

… 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.

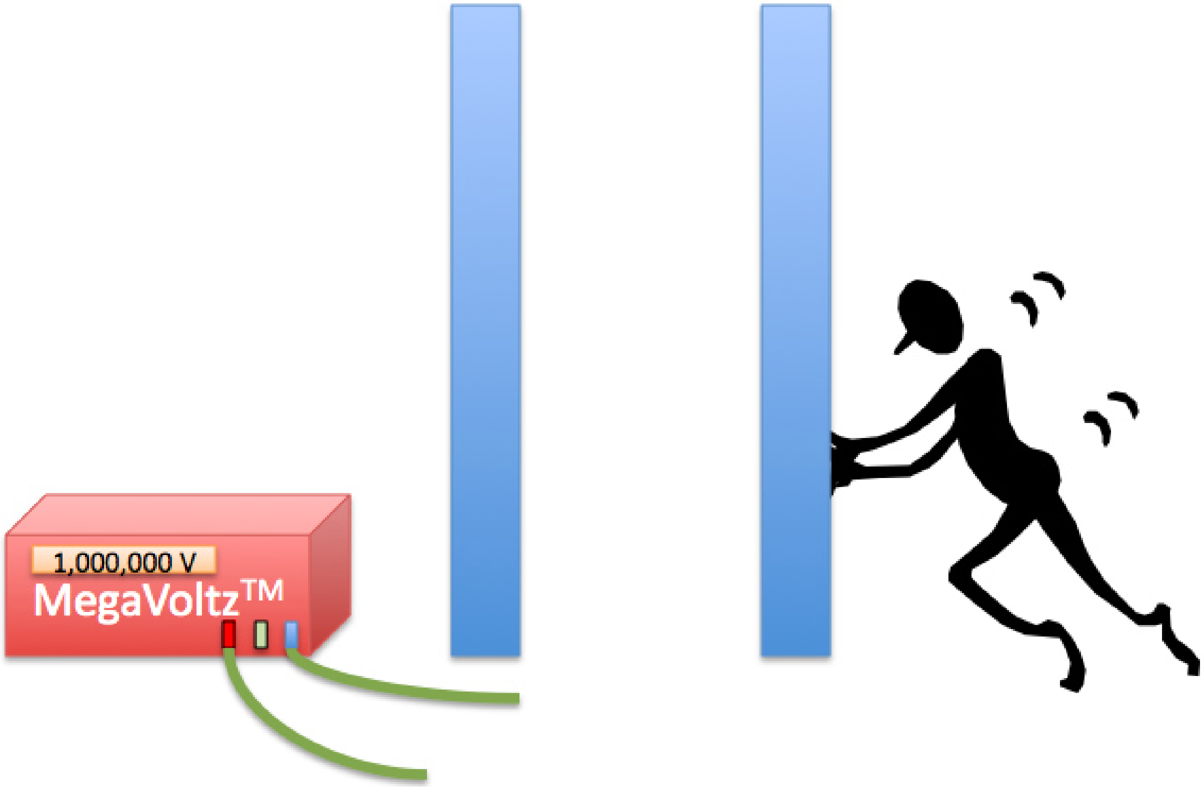
Submit Answer Tries 0

- ◯ 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.

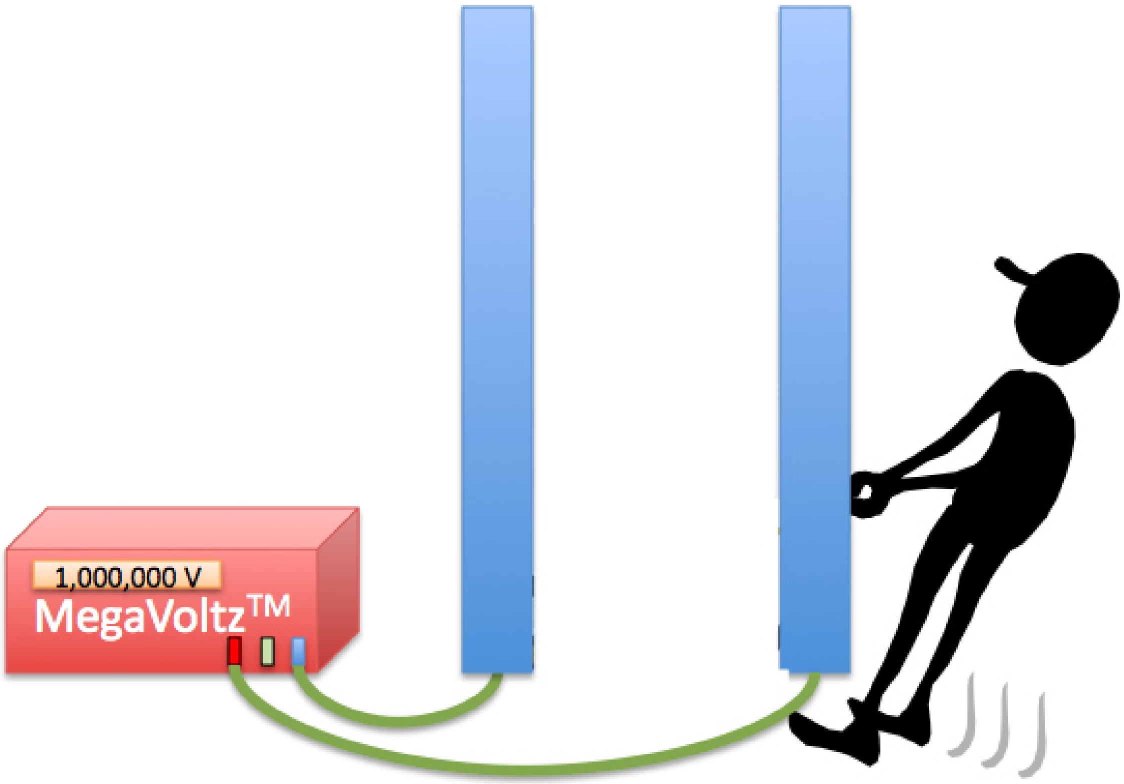
Submit Answer Tries 0

- ○ 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 …

Two ways how the paper could slide off the fridge:

- Magnet slides off paper
- Paper and magnet slide off fridge

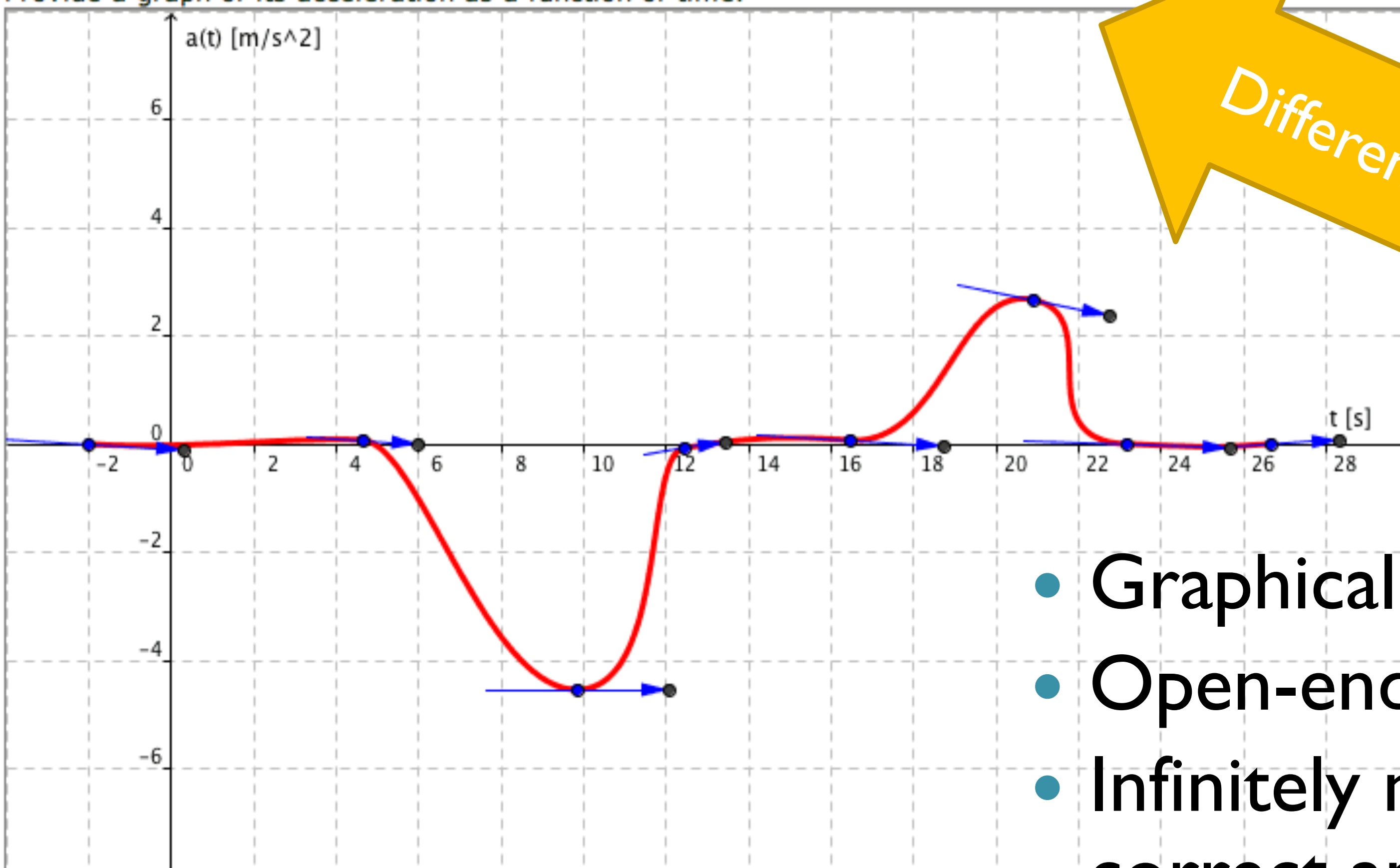
Depending on values, one or the other decides.

A sheet of paper is attached to the door of your refrigerator by a magnet. The coefficient of static friction between the fridge door and the paper is 0.6, and between the paper and the magnet is 1.4. The mass of the paper is 2 gram, the mass of the magnet is 10 gram. What is the magnitude of the minimum force with which the magnet must be attracted to the fridge, so the note sticks?

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

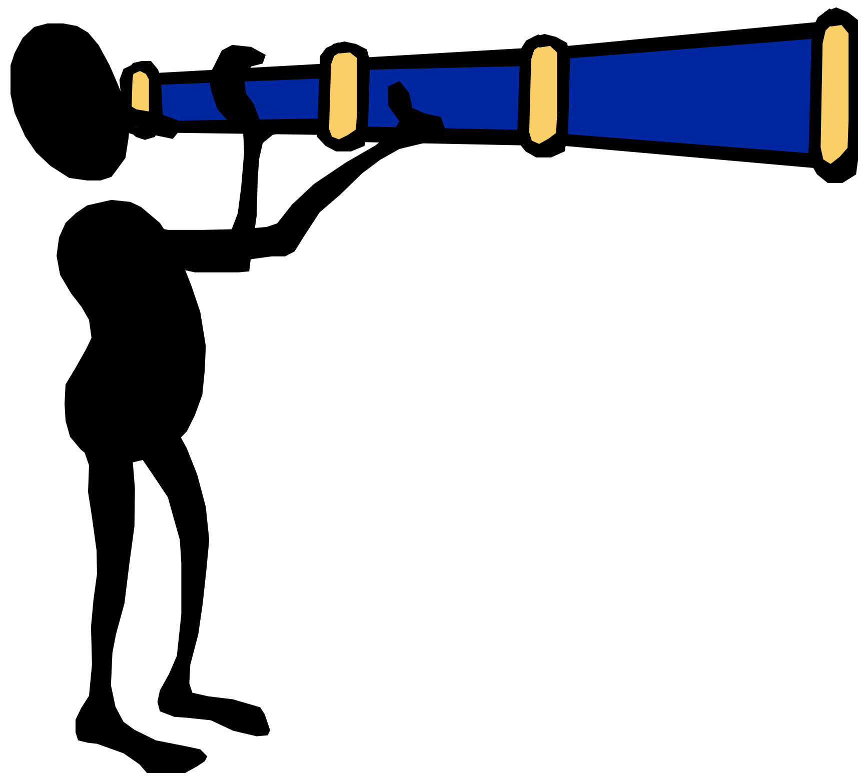
Rules Log

Different stories

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

Outlook

- More research needed how problem characteristics influence unproductive behavior



The Verdict

- Students guess and copy
  - Male students guess more,
    female students copy or collaborate more
  - High performing students guess less
  - High and low performing students copy equally much
- Success on first attempt strongly tainted by copying
  - Almost a bad sign to get it right immediately
  - Bad indicator of overall success
    - Except very early in the semester

The Verdict

- Limiting number of allowed tries to a very low number is not a good idea
  - Fosters copying or close collaboration
  - Reduces overall success on homework with no desirable effects
- Very high number is not a good idea
  - Fosters random guessing
  - Reduces overall success on homework with no desirable effects
- Five seems about right

The Verdict

- Undesirable homework behavior can be reduced by introducing more short exams
- It may be promising to have students turn in some derivations
- … or maybe give better homework.

Thank you!

- Gerd Kortemeyer
  kortemey@msu.edu