Multiple Tries on Trial

Gerd Kortemeyer
Michigan State University
The Accusation

- Online homework fosters unproductive behavior
- Too many multiple tries are at fault
  ... or maybe ...
  Too few tries are at fault
Superman Stops the Train

An out-of-control train is racing toward a city's 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

Due this Friday, Feb 27 at 11:00 pm (EST)
Typical Online Physics Problem

Superman Stop

An out-of-control train is racing toward a city. The police have shut down the streets and only Superman can help. The train is traveling at 100 km/h. Superman has a mass of 103 kg. How fast does Superman have to fly in order to stop the train in a totally inelastic steel-Man-of-Steel collision?

Multiple tries 0/5

How many?
How Many Tries to Grant?

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

![Bar chart showing distribution of maximum allowed tries for different grading policies and credit systems.](chart.png)
How Many Tries to Grant?

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

Not exactly consensus …
How Many Tries to Grant?

- Why is there no consensus?
- Balancing act

<table>
<thead>
<tr>
<th>Low Number of Allowed Tries</th>
<th>High Number of Allowed Tries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Possibly Good</td>
<td>Possibly Bad</td>
</tr>
<tr>
<td>• Better exam preparation</td>
<td>• Random guessing</td>
</tr>
<tr>
<td>• Less grade-inflation</td>
<td>• False sense of security</td>
</tr>
<tr>
<td>• Better mastery-based</td>
<td>• Encouragement</td>
</tr>
<tr>
<td>formative assessment</td>
<td>• Less whining</td>
</tr>
<tr>
<td>• Discouragement</td>
<td>• Copying</td>
</tr>
<tr>
<td>• More whining</td>
<td>• More whining</td>
</tr>
</tbody>
</table>
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?

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

Integral Submission [Percent]

Seconds between Subsequent Submissions

Guessing

Time it takes to read problem

1 min

1 hr


Male (N=85070)

Female (N=126047)
Tries versus Success

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

\[ y = 38808e^{-0.414x} \]

\[ R^2 = 0.98166 \]
Tries versus Success

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

\[
y = 962.49e^{-0.274x}
\]

\[
R^2 = 0.94869
\]
Tries versus Success

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

\[
\Delta N_s(n) = N_{s,0} \exp(-\lambda_s n)
\]
\[
\Delta N_a(n) = N_{a,0} \exp(-\lambda_a n)
\]

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

- “Probabilities” of succeeding or giving up on a particular attempt

![Graph showing the relationship between Maximum Allowed Tries and Decay Constant with a linear trend line formula y = -0.0137x + 0.6877]
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

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

### TABLE 2

<table>
<thead>
<tr>
<th>Class</th>
<th>Grade</th>
<th>Student #</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Grade &gt;= 3.5</td>
<td>69</td>
<td>30.40%</td>
</tr>
<tr>
<td>Middle</td>
<td>2.0 &lt; Grade &lt; 3.5</td>
<td>95</td>
<td>41.80%</td>
</tr>
<tr>
<td>Low</td>
<td>Grade &lt;= 2.0</td>
<td>63</td>
<td>27.80%</td>
</tr>
</tbody>
</table>

### TABLE 3

<table>
<thead>
<tr>
<th>Class</th>
<th>Grade</th>
<th>Student #</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passed</td>
<td>Grade &gt; 2.0</td>
<td>164</td>
<td>72.2%</td>
</tr>
<tr>
<td>Failed</td>
<td>Grade &lt;= 2.0</td>
<td>63</td>
<td>27.80%</td>
</tr>
</tbody>
</table>
**Data Mining Access Logs**

- See how much you can explain

<table>
<thead>
<tr>
<th>Classifier</th>
<th>2-Classes</th>
<th>3-Classes</th>
<th>9-Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tree Classifier</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5.0</td>
<td>80.3</td>
<td>56.8</td>
<td>25.6</td>
</tr>
<tr>
<td>CART</td>
<td>81.5</td>
<td>59.9</td>
<td>33.1</td>
</tr>
<tr>
<td>QUEST</td>
<td>80.5</td>
<td>57.1</td>
<td>20.0</td>
</tr>
<tr>
<td>CRUISE</td>
<td>81.0</td>
<td>54.9</td>
<td>22.9</td>
</tr>
<tr>
<td><strong>Non-tree Classifier</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayes</td>
<td>76.4</td>
<td>48.6</td>
<td>23.0</td>
</tr>
<tr>
<td>1NN</td>
<td>76.8</td>
<td>50.5</td>
<td>29.0</td>
</tr>
<tr>
<td>kNN</td>
<td>82.3</td>
<td>50.4</td>
<td>28.5</td>
</tr>
<tr>
<td>Parzen</td>
<td>75.0</td>
<td>48.1</td>
<td>21.5</td>
</tr>
<tr>
<td>MLP</td>
<td>79.5</td>
<td>50.9</td>
<td>-</td>
</tr>
<tr>
<td>CMC</td>
<td>86.8</td>
<td>70.9</td>
<td>51.0</td>
</tr>
</tbody>
</table>
Data Mining Access Logs

... and find the most important features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Importance %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total_Correct_Answers</td>
<td>100.00</td>
</tr>
<tr>
<td>Total_Number_of_Tries</td>
<td>58.61</td>
</tr>
<tr>
<td>First_Got_Correct</td>
<td>27.70</td>
</tr>
<tr>
<td>Time_Spent_to_Solve</td>
<td>24.60</td>
</tr>
<tr>
<td>Total_Time_Spent</td>
<td>24.47</td>
</tr>
<tr>
<td>Communication</td>
<td>9.21</td>
</tr>
</tbody>
</table>
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!

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!

Item Response Theory

- Modeling unproductive behavior
- Need new IRT model

\[ \hat{\pi}_{ij} = \chi_j \left(1 - p_{ij}\right) + \left(1 - \gamma_j\right)p_{ij} \]

\[ = \chi_j + \frac{1 - \gamma_j - \chi_j}{1 + \exp\left(a_i(b_i - \theta_j)\right)} \]

- Guessing and copying as learner traits
Item Response Theory

- 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.
Why?

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

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?
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
More Exams

- Sanctioned internal discussions, where course instructors participate

More Exams

More Exams

- The proof is in the pudding: Final Exam

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
A car (mass of 750 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 speeding up with 2.3 m/s². What is 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/s². 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/s². What is the magnitude of the lifting force?

Lifting/lowering, speeding up/slowing down, different numbers
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
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
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/s².

Provide a graph of its acceleration as a function of time.

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

- More research needed on 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