Developing Systems that Detect and Adapt to When Students Game the System

Ryan Shaun Joazeiro de Baker
Learning Sciences Research Institute
University of Nottingham
ryan@educationaldatamining.org

Outline

- Introduction

...

Introduction

- Today I’ll be talking about detecting and responding to student behavior within a specific type of interactive learning environment, Cognitive Tutors

Cognitive Tutors

- One of the most widely used AI-based educational technologies

- Developed for a variety of domains (Geometry, Algebra I&II, LISP, etc)
  (Anderson et al, 1995)

- Now used in math courses in 6% of USA middle schools and high schools (marketed by Carnegie Learning, Inc.)

Carnegie Learning®

Cognitive Tutors

- Student completes math problems
- Cognitive Tutor gives feedback & help based on task analysis of skills being taught, and model of student cognition
- Student is assigned new problems based on what skills they have not yet mastered

Cognitive Tutors

- Are effective (~1 SD better than traditional classroom instruction)

- Still not as good as expert human tutors (2 SD)

  (Bloom 1984)

- And routinely, a small percentage of students learn very little

Clip of Student Behavior

- I'd like to show you how one student chose to use an intelligent tutor

- ~13 year old female student
- Suburbs of a city in the Northeast USA

- Using a tutor known to be effective
  - Average student achieves 51% of potential pre-post gain

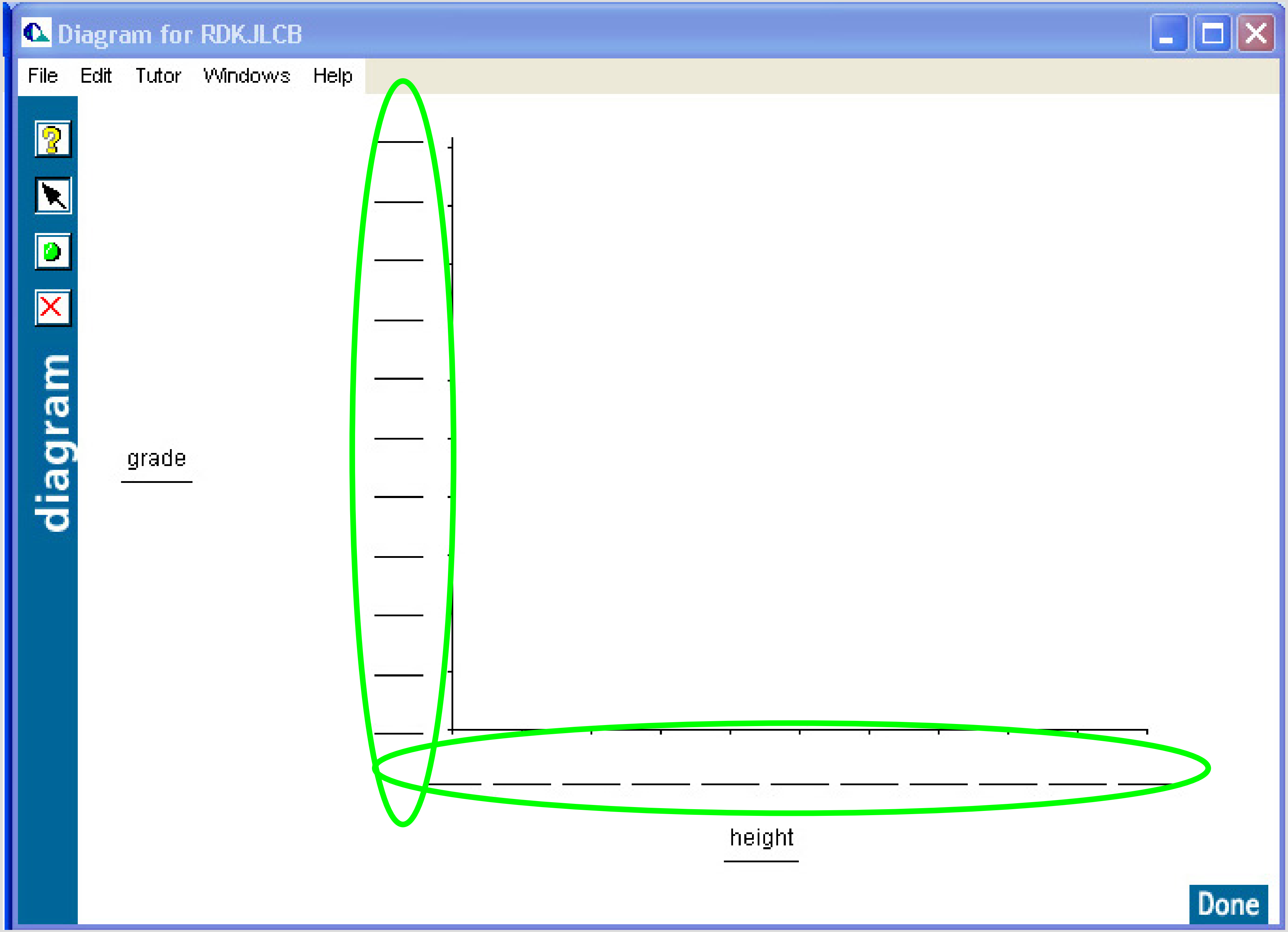
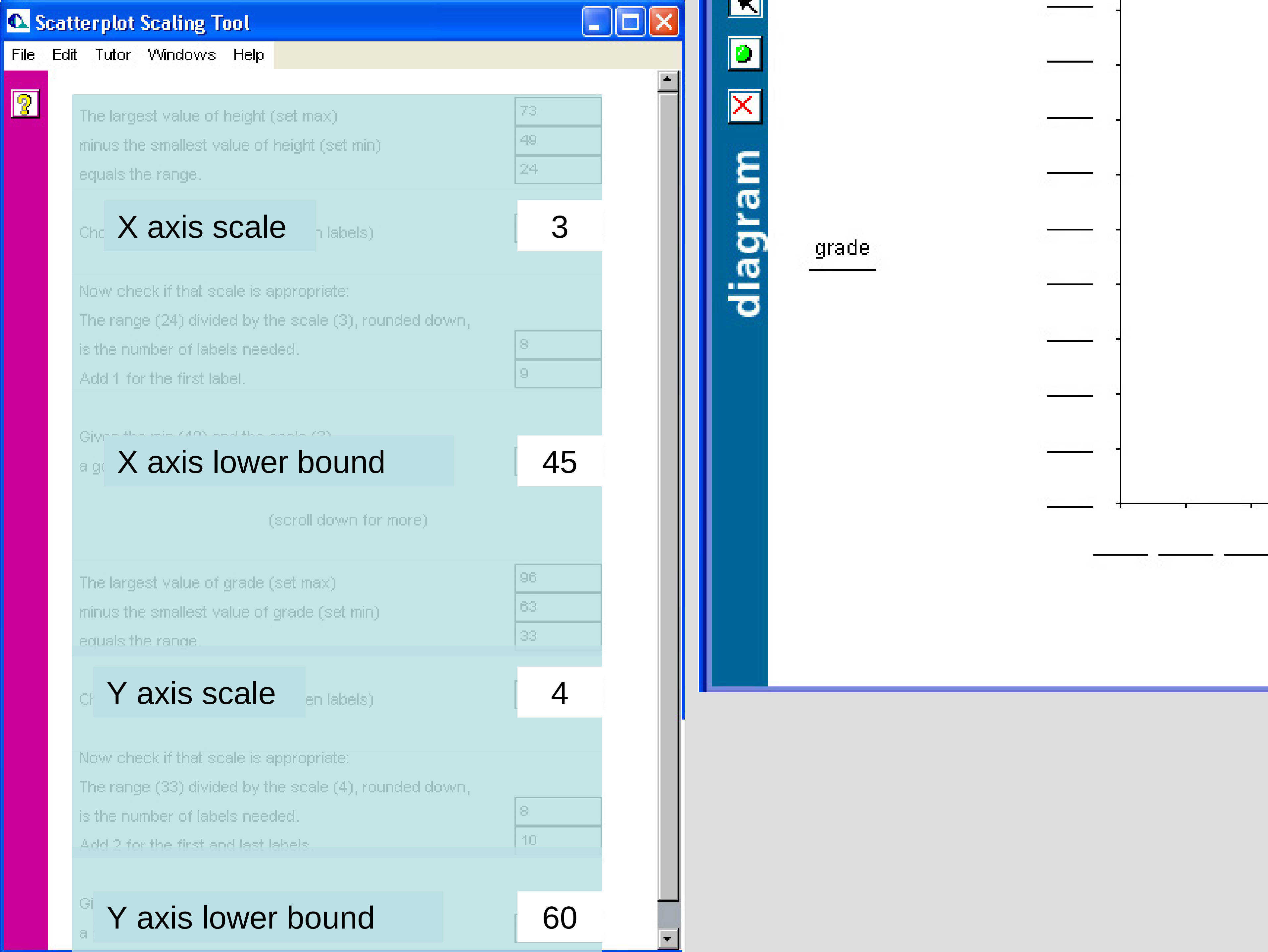
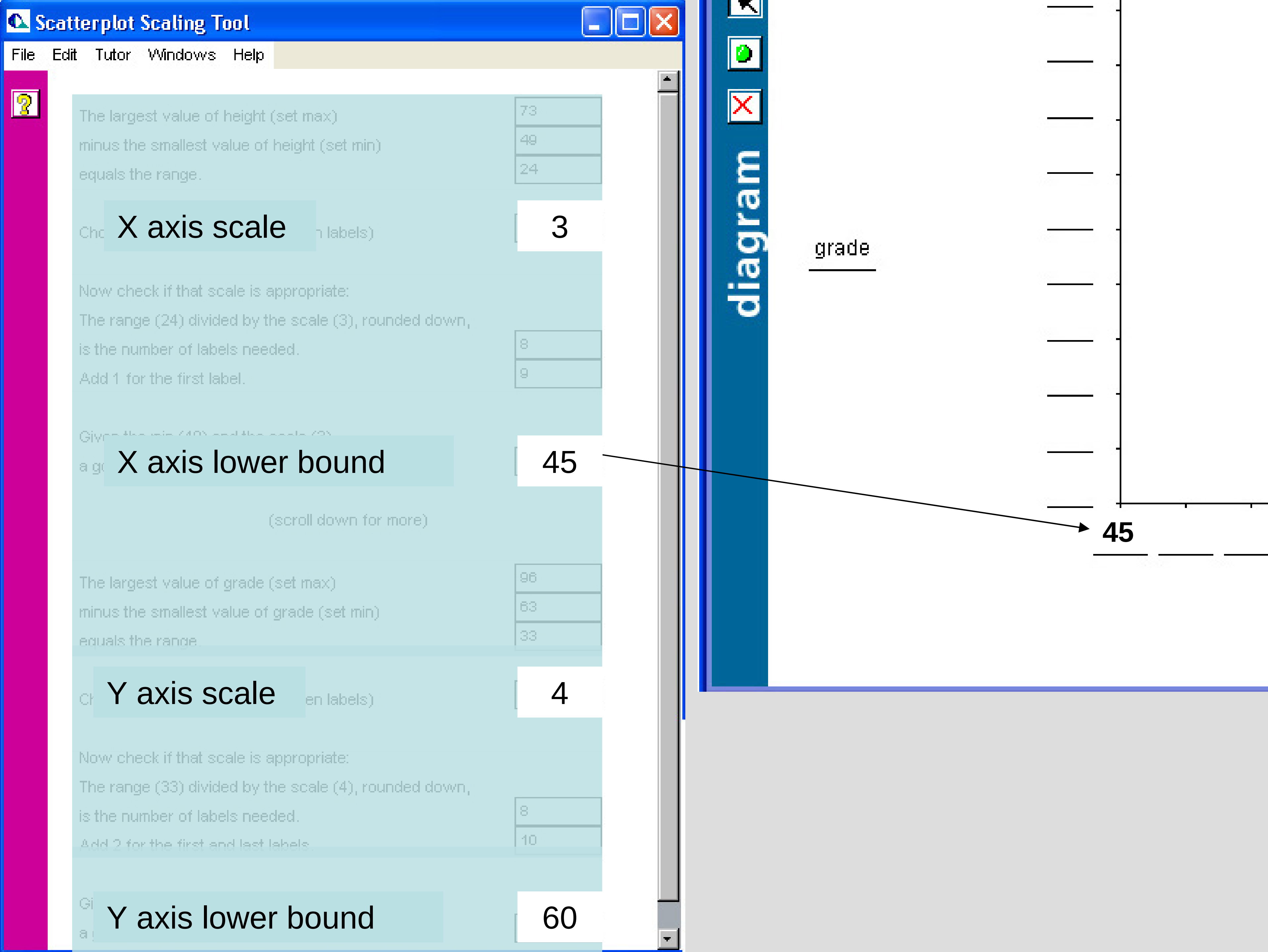
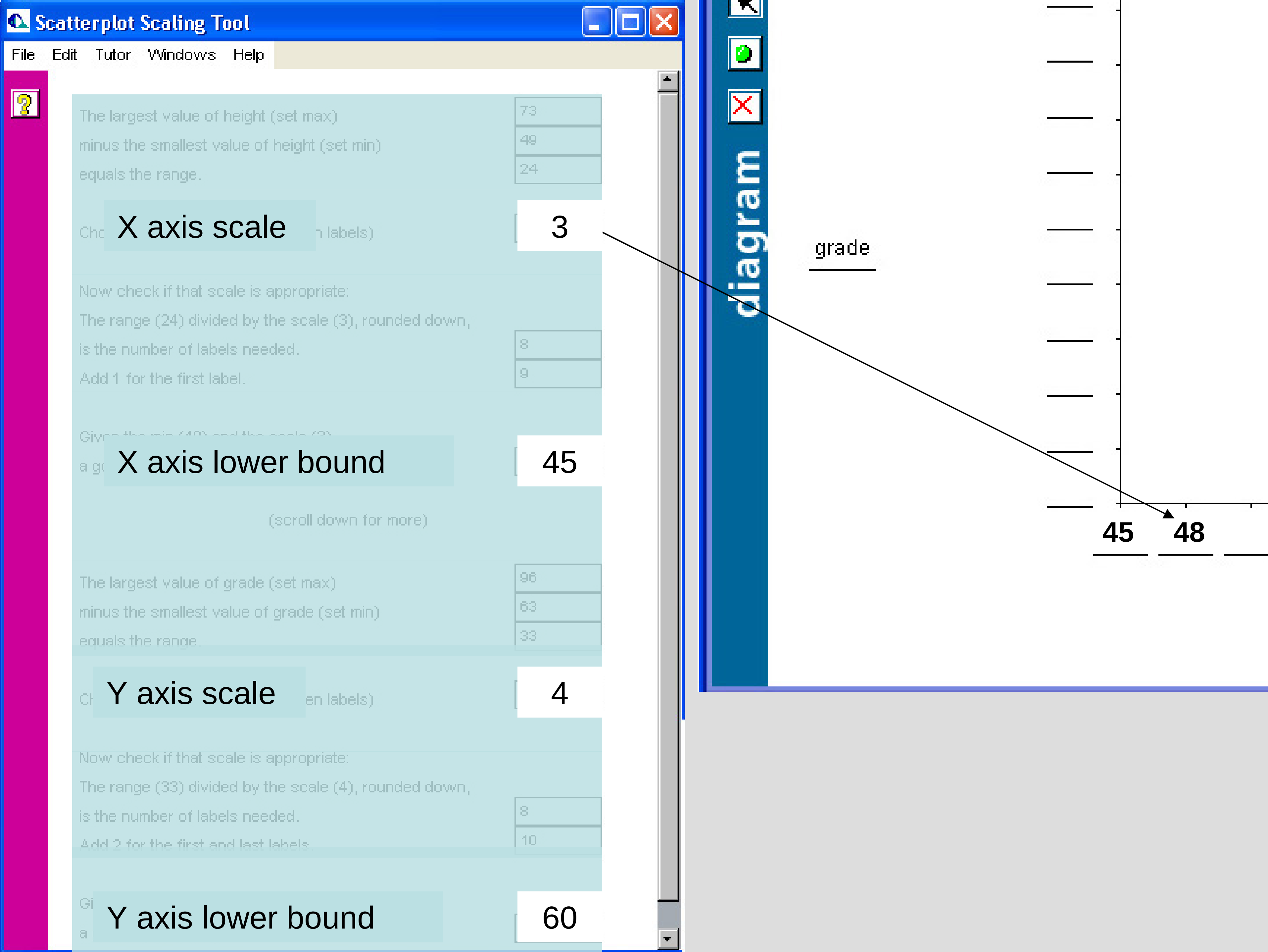


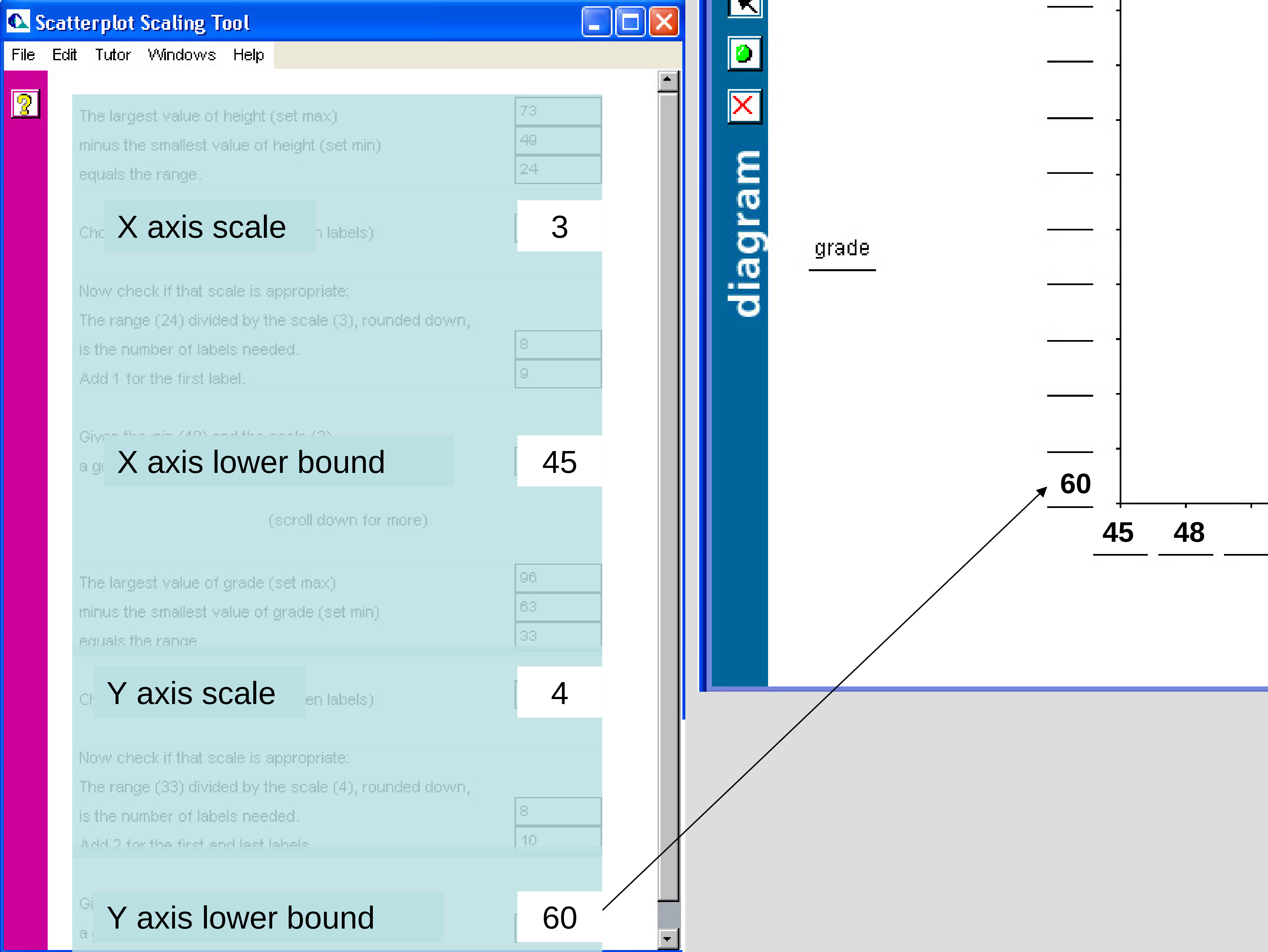
Diagram for RDKJLCB
File
Edit
Tutor
Windows
Help
diagram
grade
height
Done



Scatterplot Scaling Tool
File Edit Tutor Windows Help
The largest value of height (set max) 73
minus the smallest value of height (set min) 49
equals the range. 24
X axis scale 3
Now check if that scale is appropriate:
The range (24) divided by the scale (3), rounded down,
is the number of labels needed. 8
Add 1 for the first label. 9
X axis lower bound 45
(scroll down for more)
The largest value of grade (set max) 96
minus the smallest value of grade (set min) 63
equals the range. 33
Y axis scale 4
Now check if that scale is appropriate:
The range (33) divided by the scale (4), rounded down,
is the number of labels needed. 8
Add 2 for the first and last labels. 10
Y axis lower bound 60
diagram
grade


Scatterplot Scaling Tool
File Edit Tutor Windows Help
The largest value of height (set max) 73
minus the smallest value of height (set min) 49
equals the range. 24
X axis scale 3
Now check if that scale is appropriate:
The range (24) divided by the scale (3), rounded down,
is the number of labels needed. 8
Add 1 for the first label. 9
X axis lower bound 45
(scroll down for more)
The largest value of grade (set max) 96
minus the smallest value of grade (set min) 63
equals the range. 33
Y axis scale 4
Now check if that scale is appropriate:
The range (33) divided by the scale (4), rounded down,
is the number of labels needed. 8
Add 2 for the first and last labels. 10
Y axis lower bound 60
diagram
grade
45


Scatterplot Scaling Tool
File Edit Tutor Windows Help
The largest value of height (set max)
73
minus the smallest value of height (set min)
49
equals the range.
24
X axis scale
3
labels)
Now check if that scale is appropriate:
The range (24) divided by the scale (3), rounded down,
is the number of labels needed.
8
Add 1 for the first label.
9
X axis lower bound
45
(scroll down for more)
The largest value of grade (set max)
96
minus the smallest value of grade (set min)
63
equals the range
33
Y axis scale
4
en labels)
Now check if that scale is appropriate:
The range (33) divided by the scale (4), rounded down,
is the number of labels needed.
8
Add 2 for the first and last labels
10
Y axis lower bound
60
diagram
grade
45
48


Scatterplot Scaling Tool
File Edit Tutor Windows Help
The largest value of height (set max)
73
minus the smallest value of height (set min)
49
equals the range.
24
X axis scale
3
labels)
Now check if that scale is appropriate:
The range (24) divided by the scale (3), rounded down,
is the number of labels needed.
8
Add 1 for the first label.
9
X axis lower bound
45
(scroll down for more)
The largest value of grade (set max)
96
minus the smallest value of grade (set min)
63
equals the range.
33
Y axis scale
4
en labels)
Now check if that scale is appropriate:
The range (33) divided by the scale (4), rounded down,
is the number of labels needed.
8
Add 2 for the first and last labels.
10
Y axis lower bound
60
diagram
grade
60
45
48

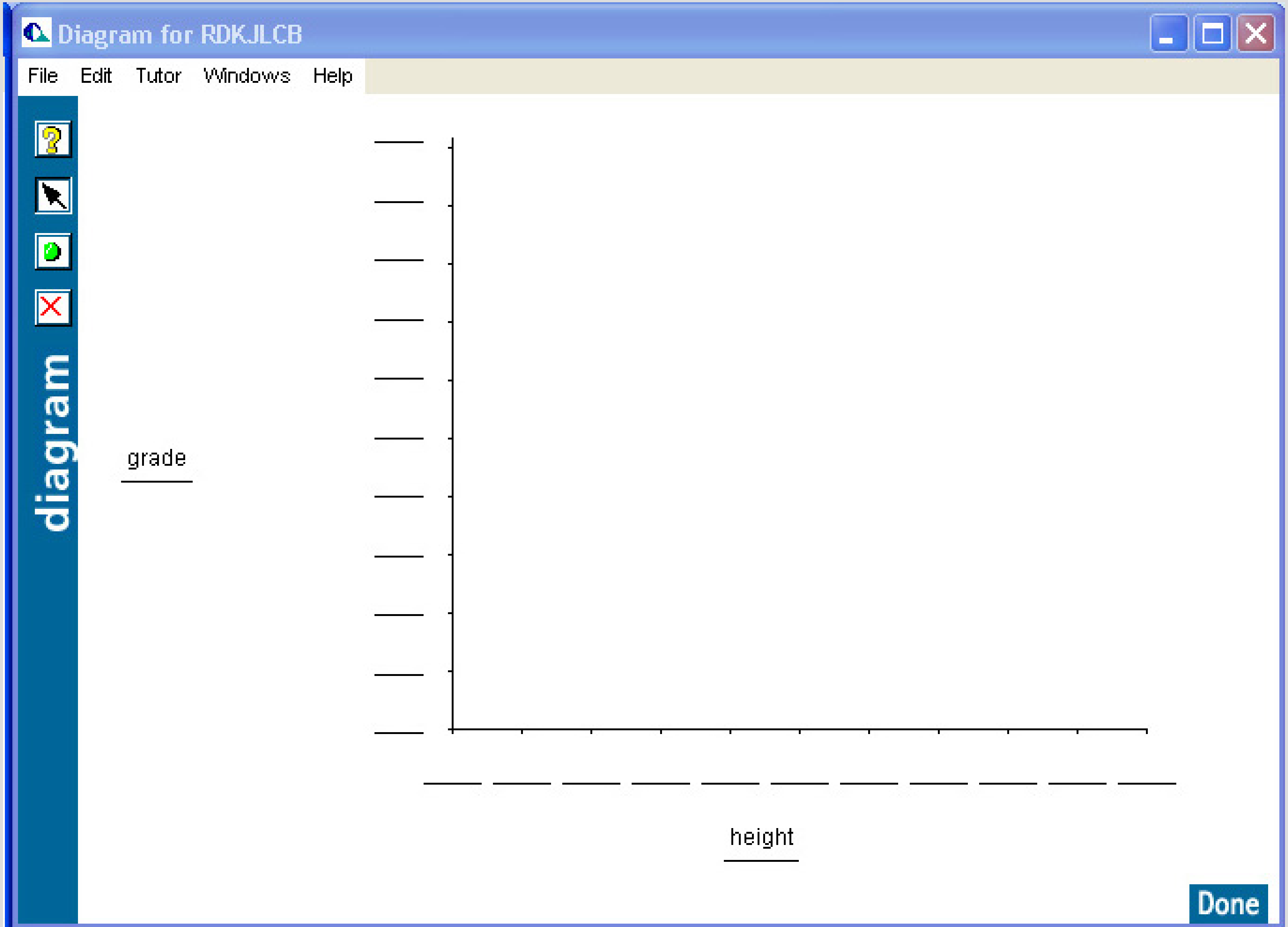


Diagram for RDKJLCB
File Edit Tutor Windows Help
diagram
grade
height
Done

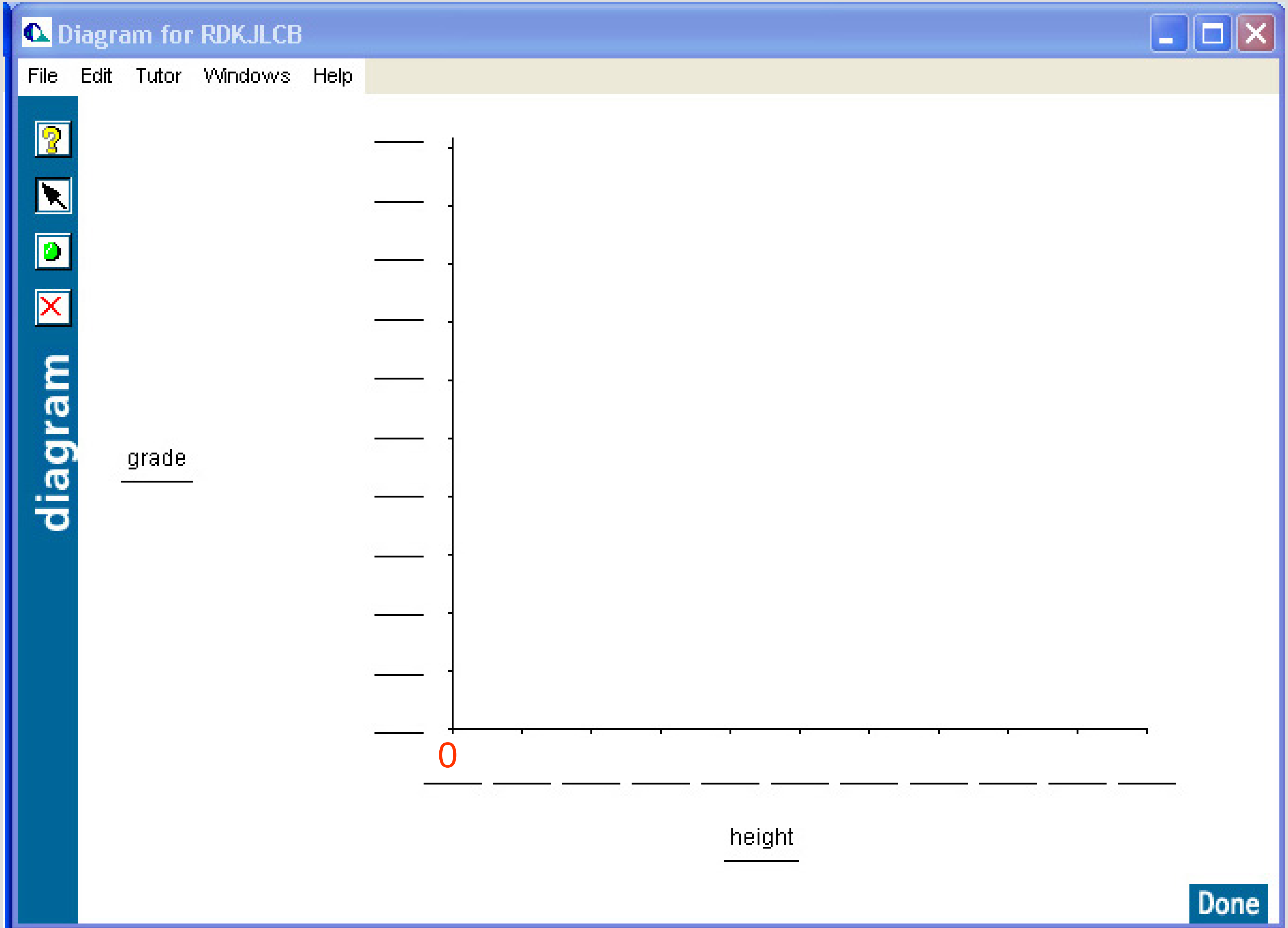


Diagram for RDKJLCB
File
Edit
Tutor
Windows
Help
diagram
grade
0
height
Done

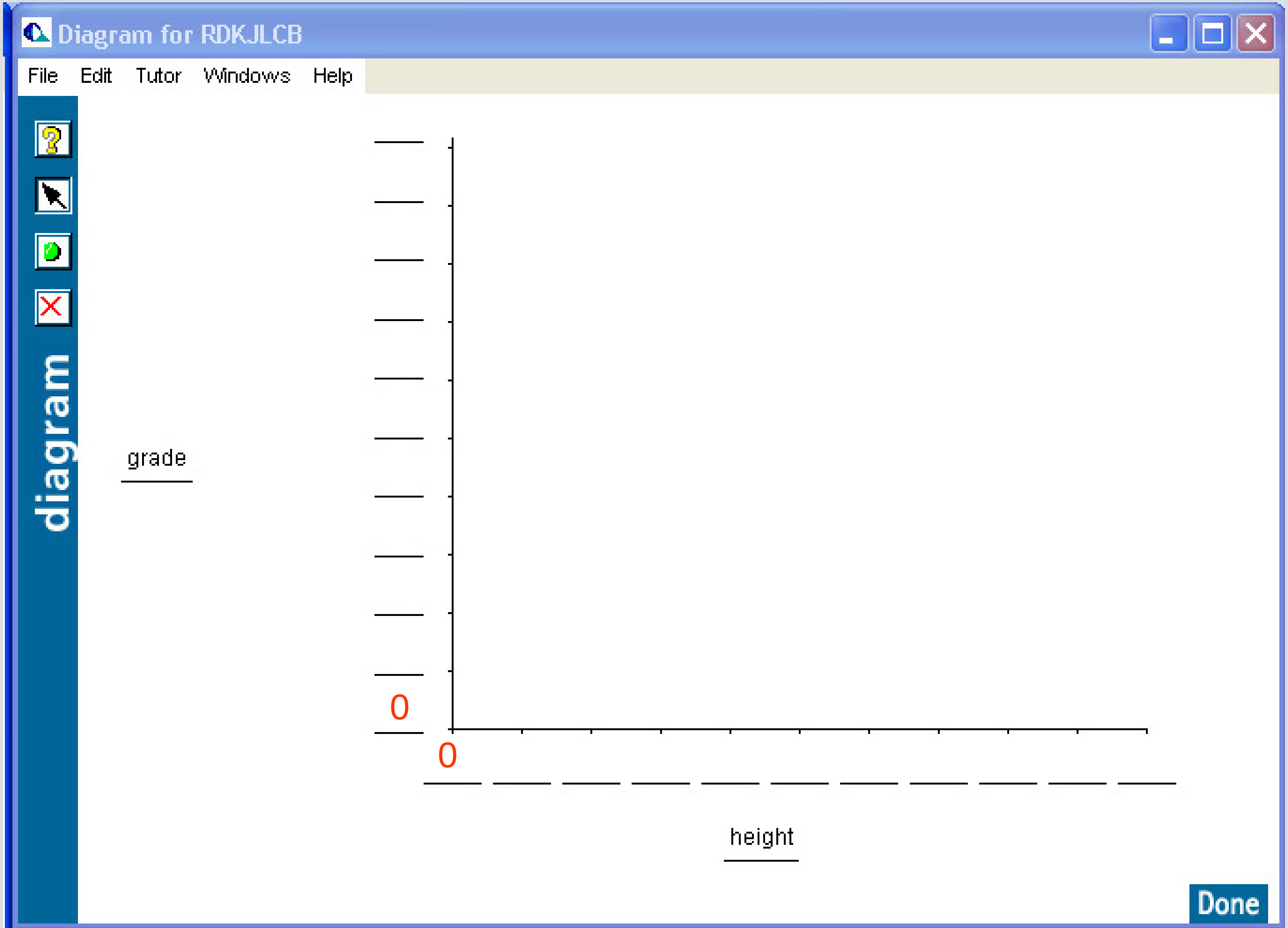


Diagram for RDKJLCB
File
Edit
Tutor
Windows
Help
diagram
grade
0
0
height
Done

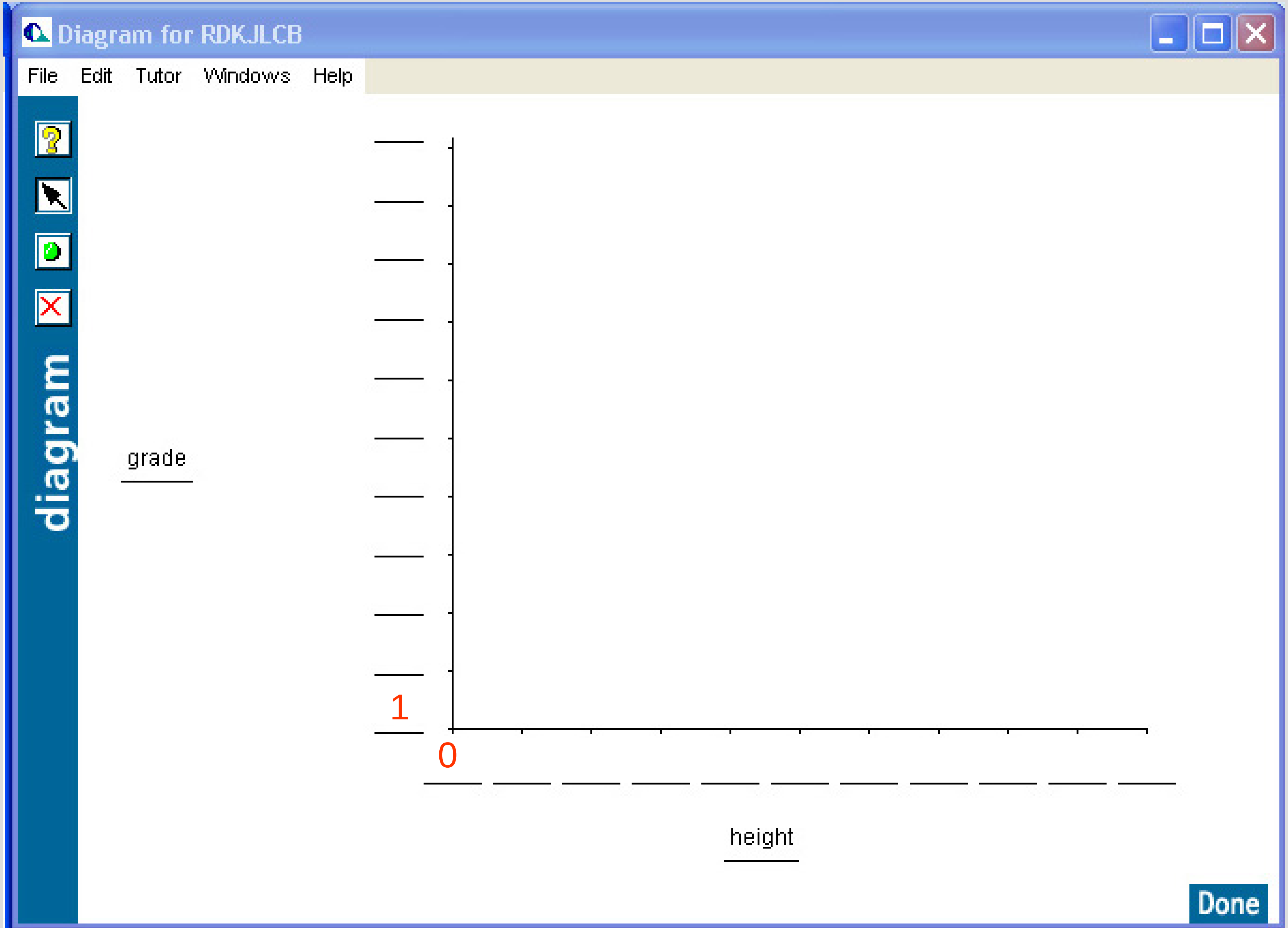


Diagram for RDKJLCB
File Edit Tutor Windows Help
diagram
grade
1
0
height
Done

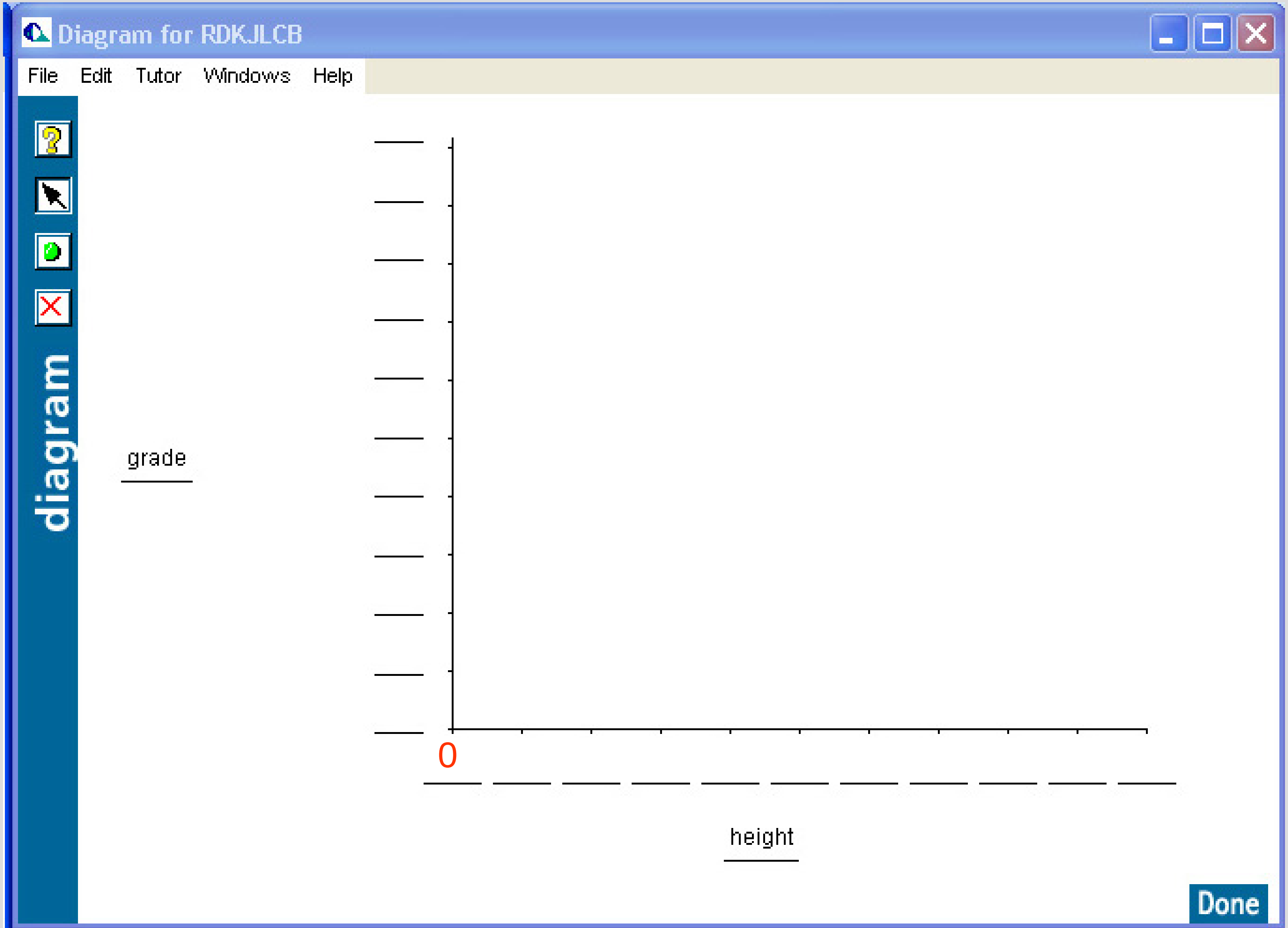


Diagram for RDKJLCB
File
Edit
Tutor
Windows
Help
diagram
grade
0
height
Done

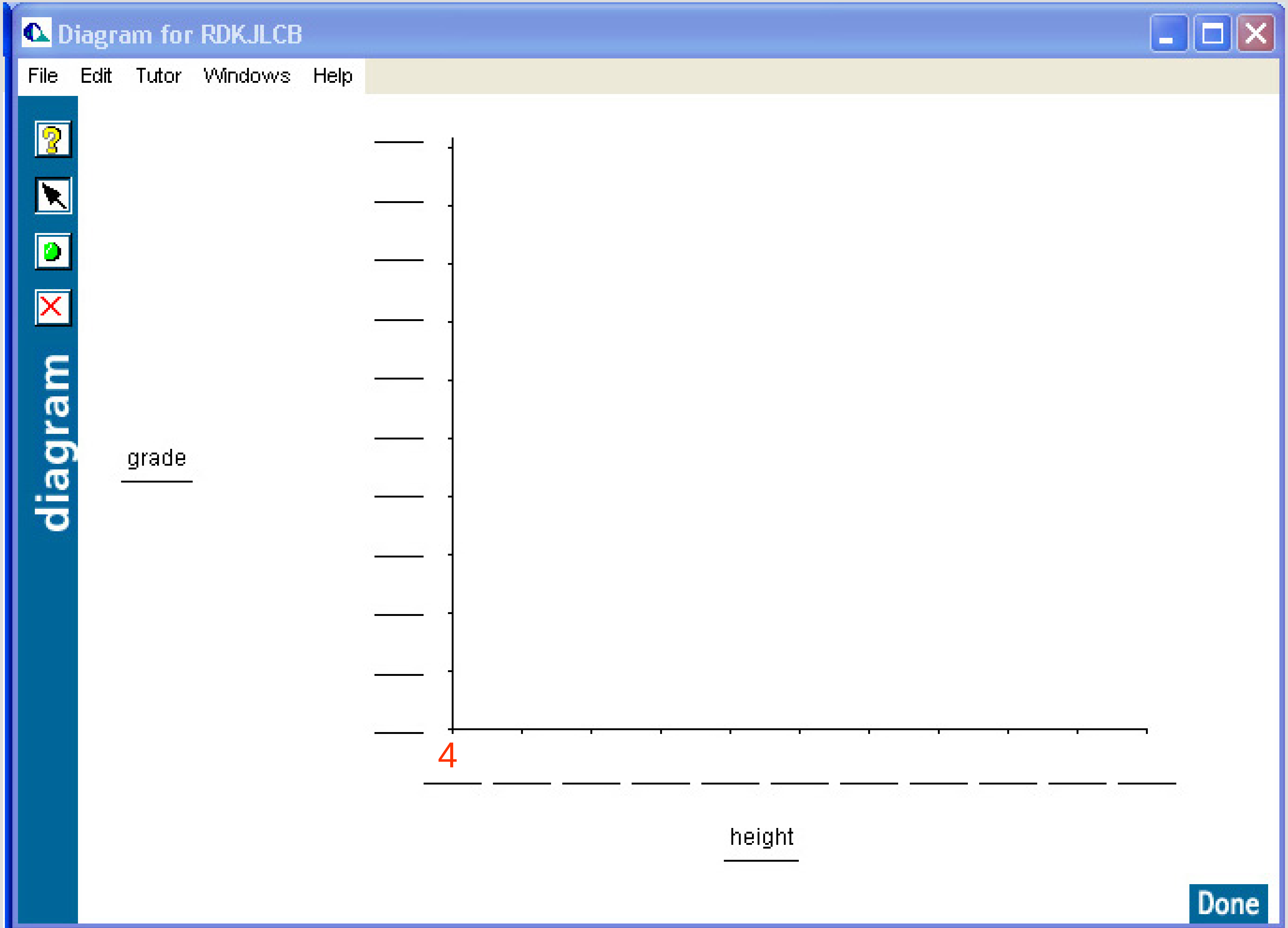
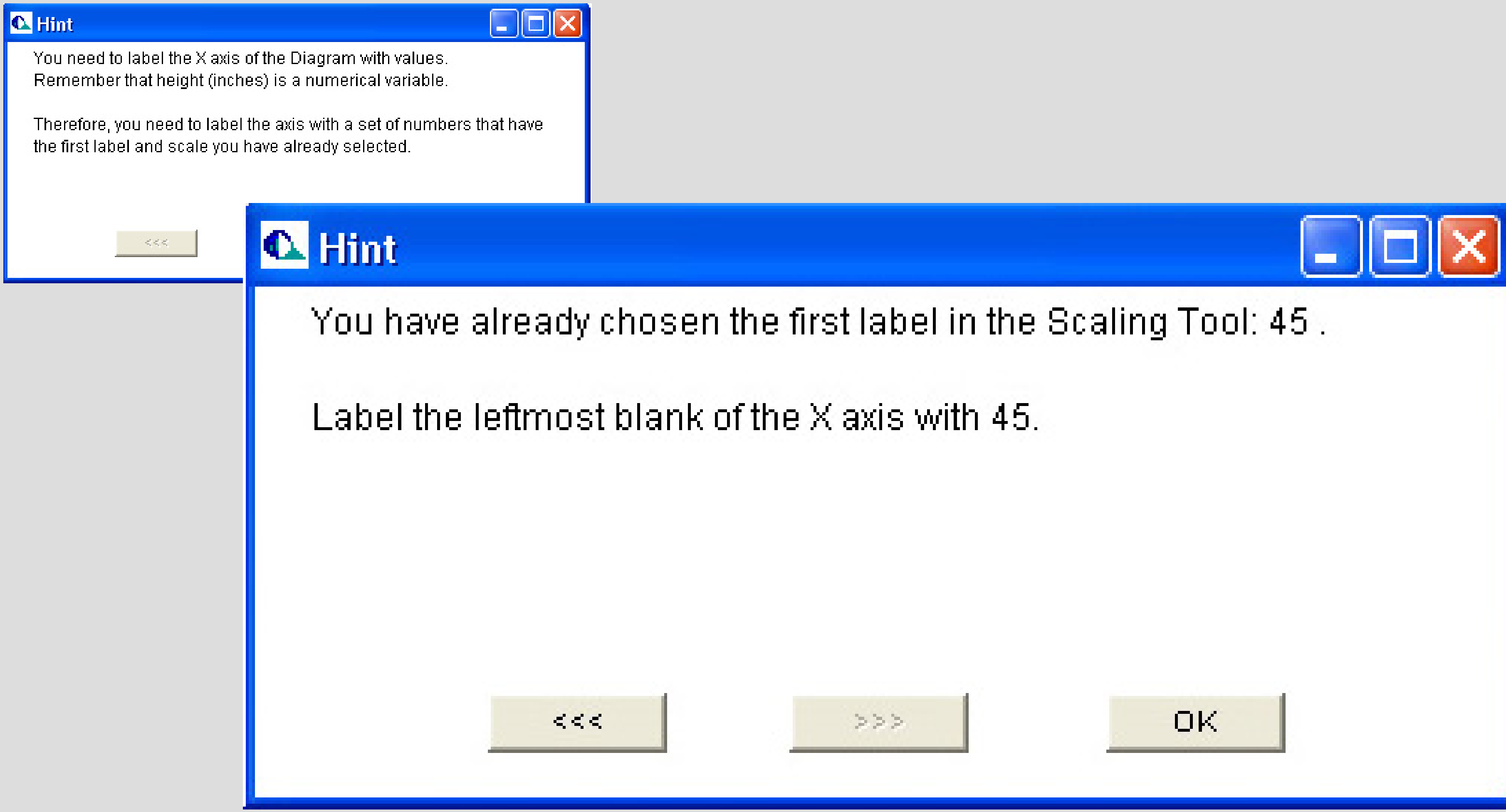


Diagram for RDKJLCB
File
Edit
Tutor
Windows
Help
diagram
grade
4
height
Done

4.8 seconds to "read" 2 levels of help and type in next answer (~1,000 wpm)



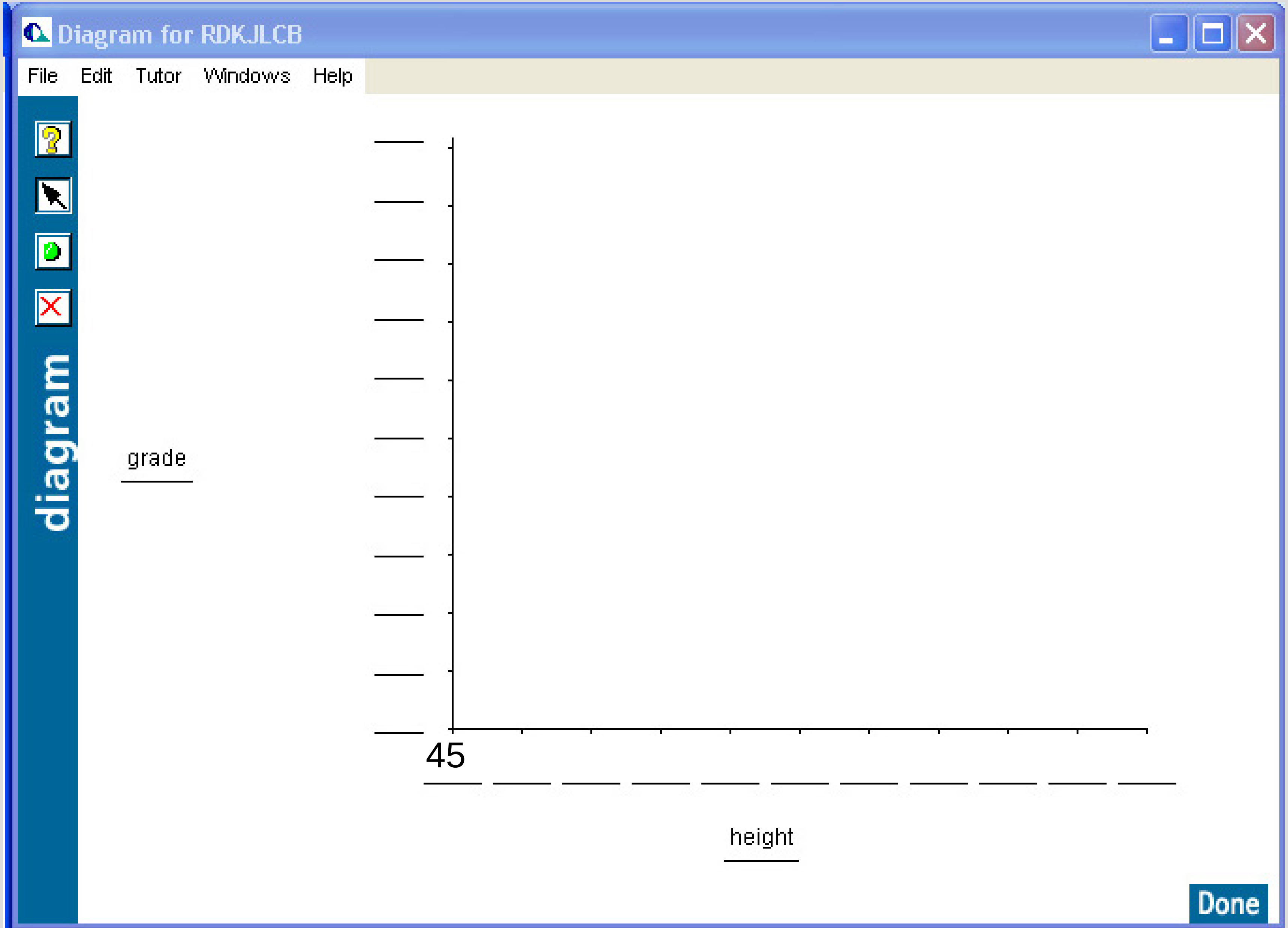


Diagram for RDKJLCB
File
Edit
Tutor
Windows
Help
diagram
grade
45
height
Done

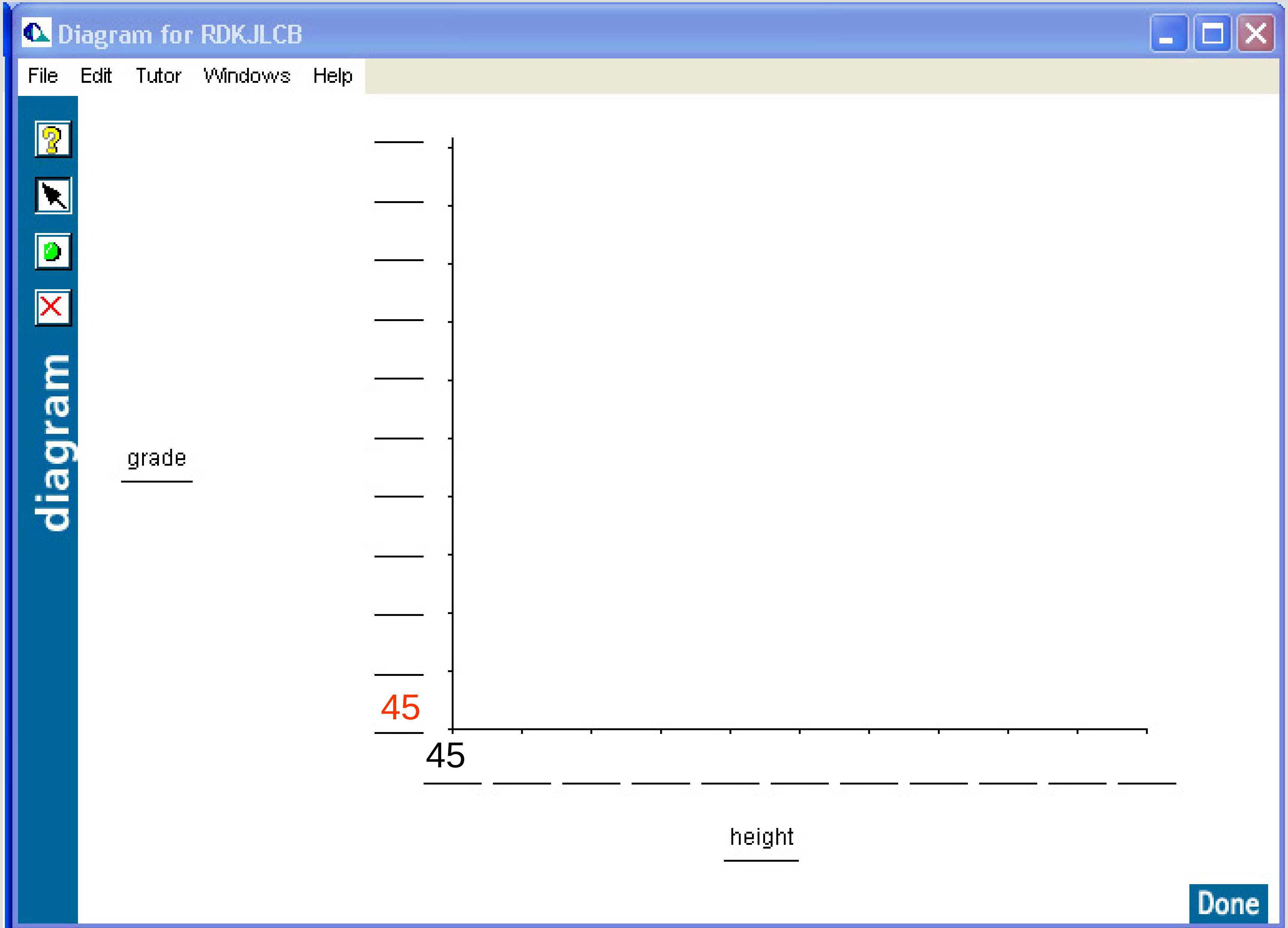
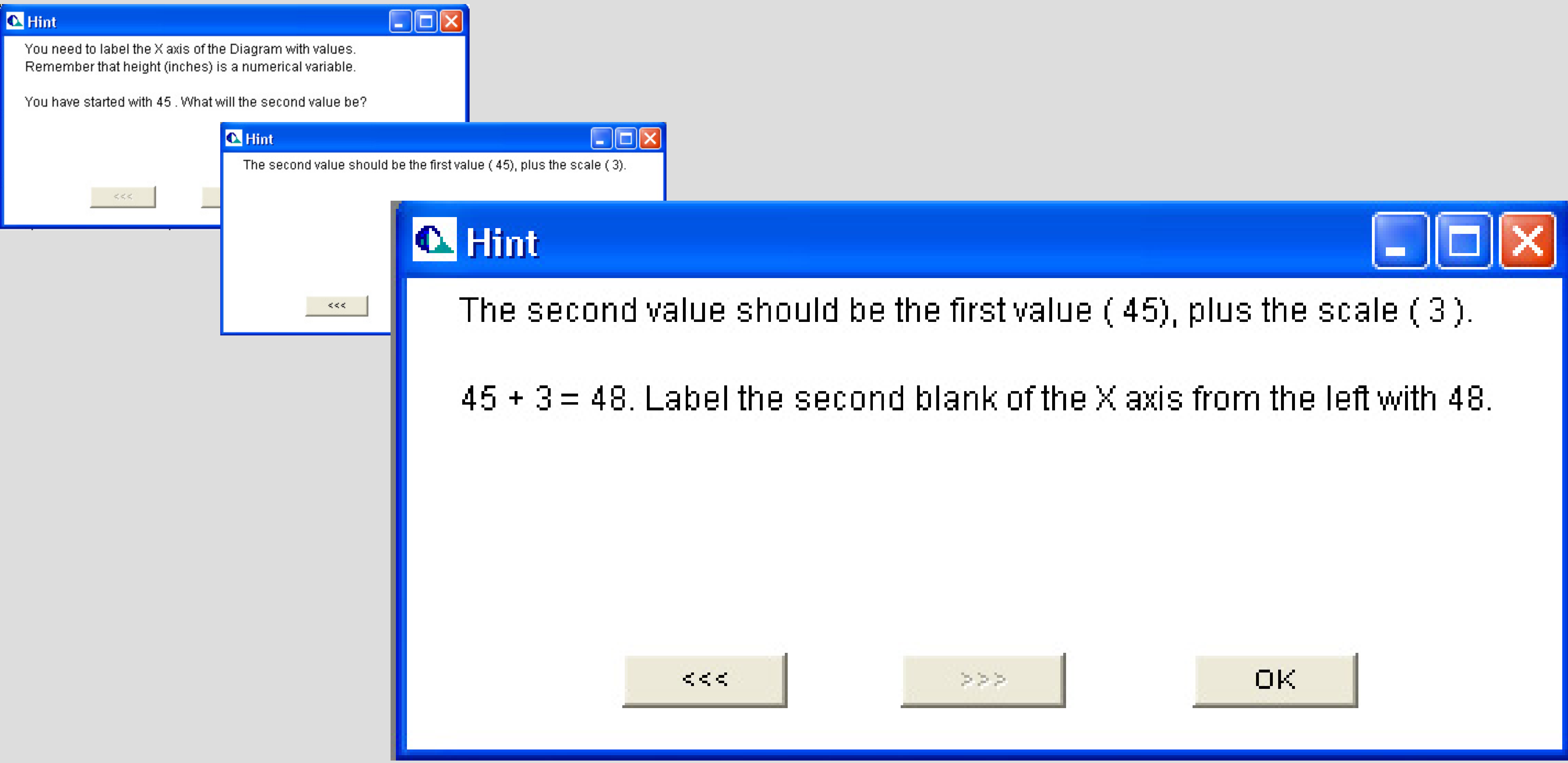


Diagram for RDKJLCB
File
Edit
Tutor
Windows
Help
diagram
grade
45
45
height
Done

5.6 seconds to “read” 3 levels of help and type in next answer (~750 wpm)



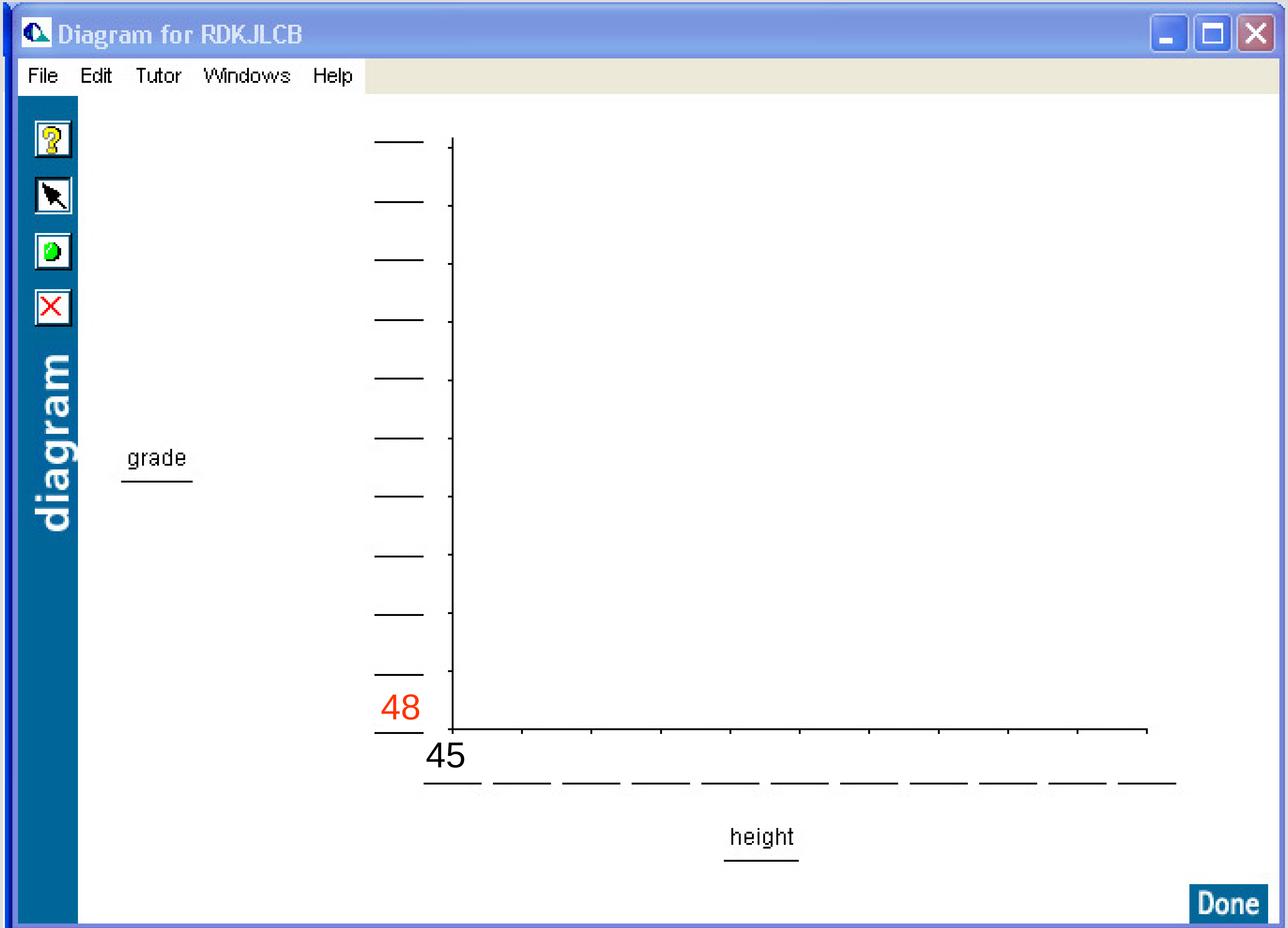
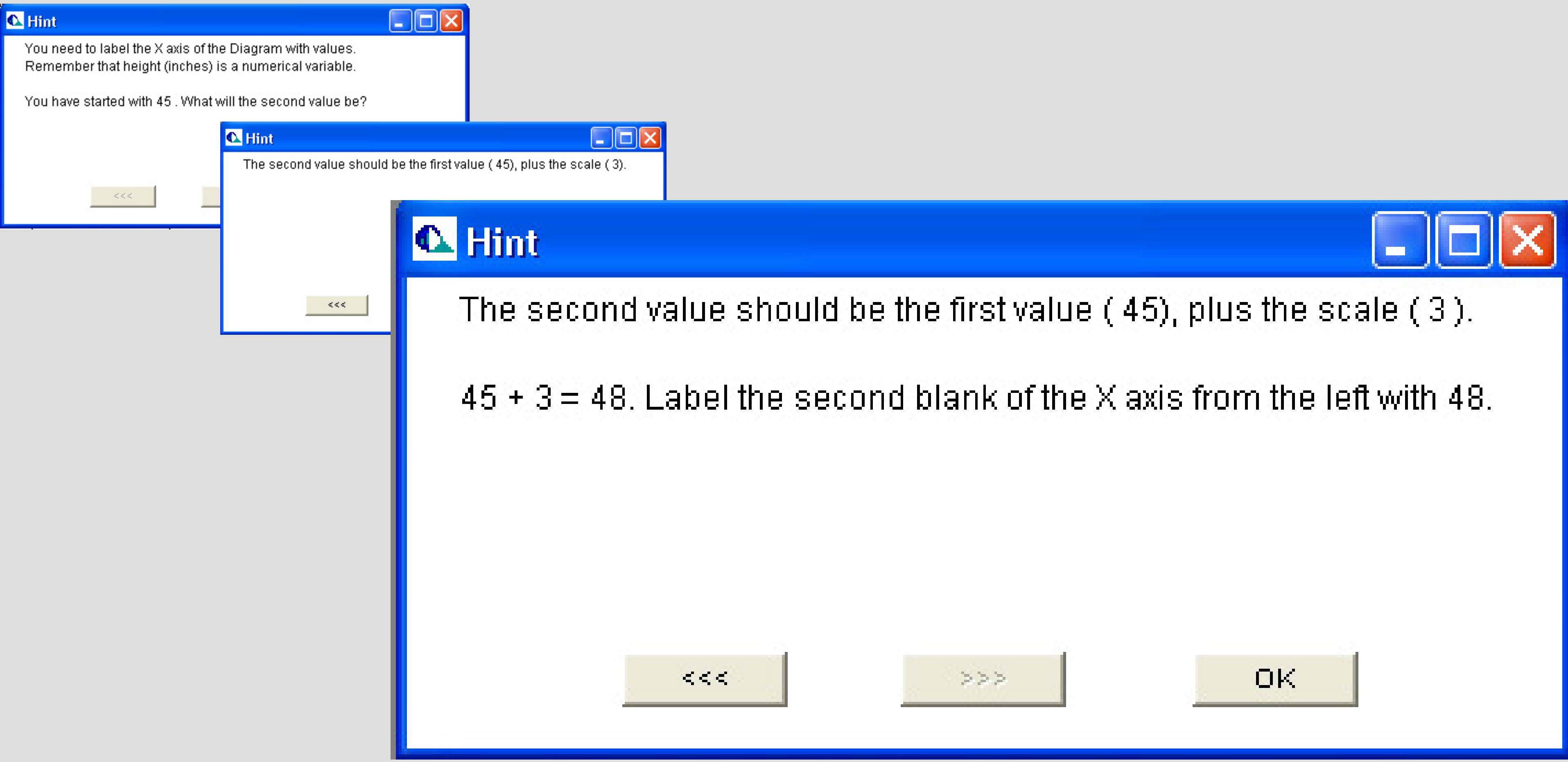


Diagram for RDKJLCB
File
Edit
Tutor
Windows
Help
diagram
grade
48
45
height
Done

6.5 seconds to reread the same help and type in next answer



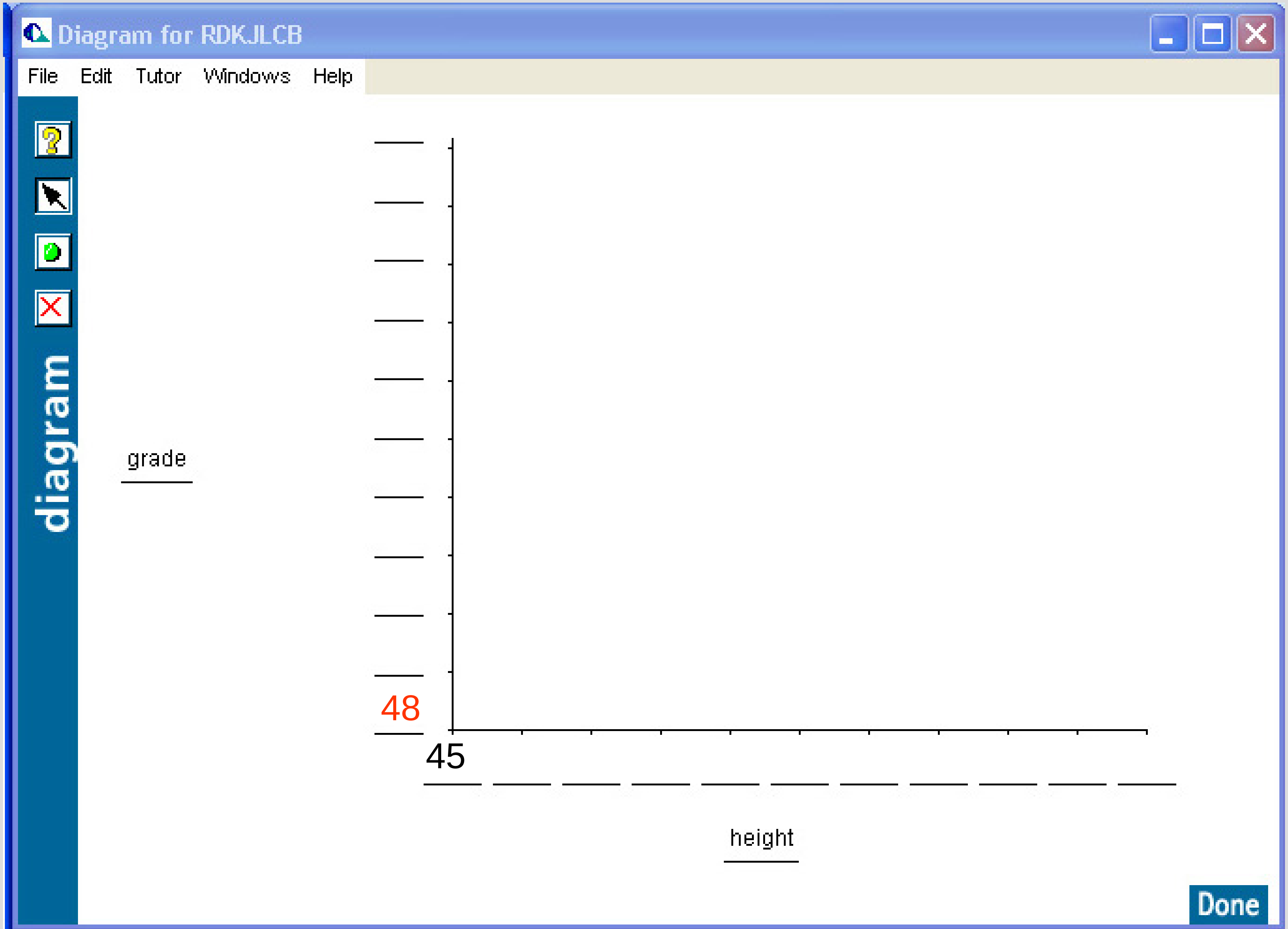
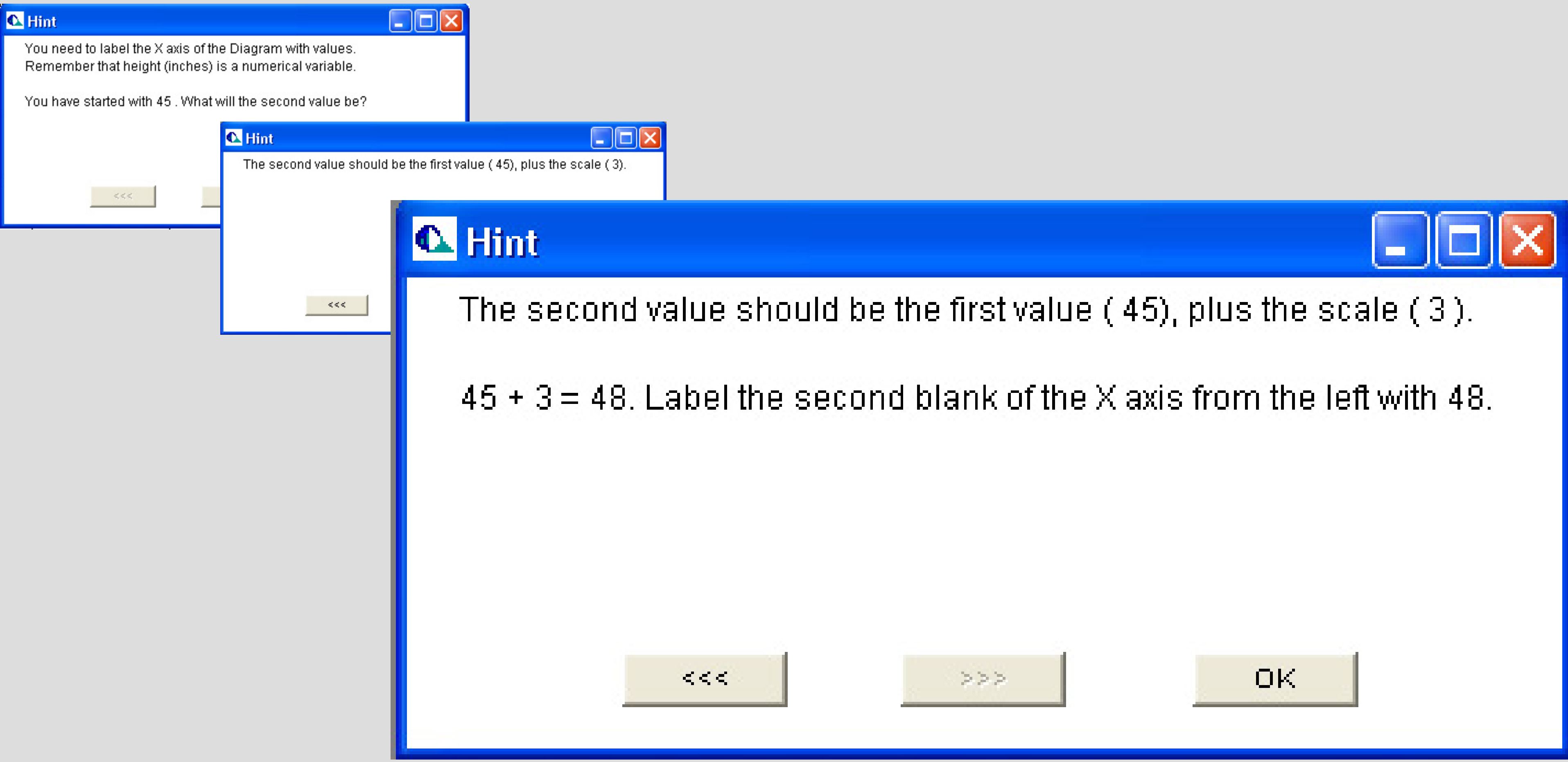


Diagram for RDKJLCB
File
Edit
Tutor
Windows
Help
diagram
grade
48
45
height
Done

4.6 seconds to reread the same help and type in next answer



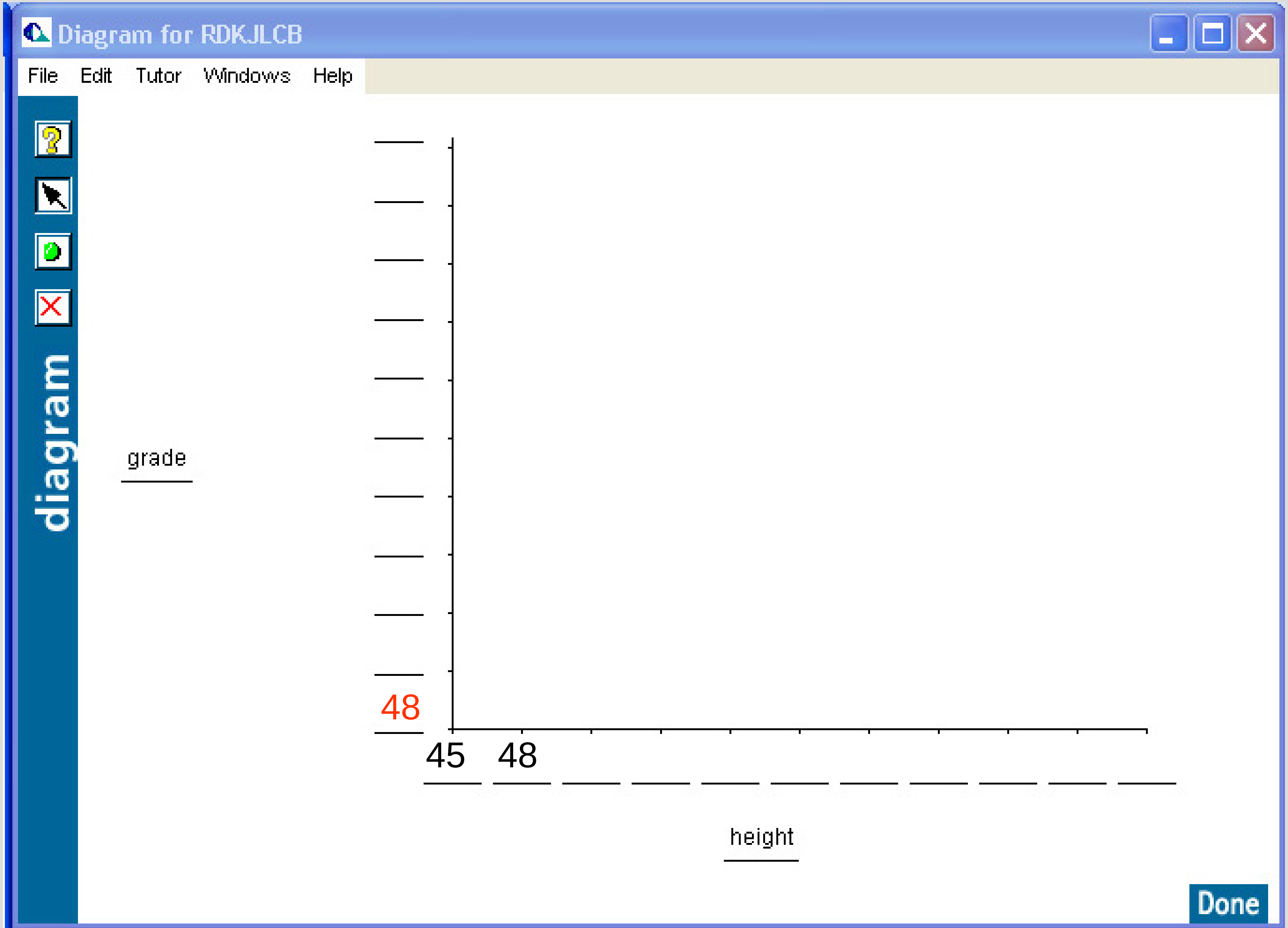


Diagram for RDKJLCB
File
Edit
Tutor
Windows
Help
diagram
grade
48
45
48
height
Done

Student Learning

- The clip you saw…
- Who thinks this student learned a lot from using the tutor?

In Fact…

- The student got 0% on the pre-test

In Fact…

- The student got 0% on the pre-test

- Then she completed 5 problems in the tutor (average = 4.14 problems)

In Fact…

- The student got 0% on the pre-test
- Then she completed 5 problems in the tutor (average = 4.14 problems)
- Then she got 0% on the post-test

Gaming the System

- The clip you saw was an instance of a behavior we call “gaming the system”

“Attempting to get correct answers and advance in a curriculum by taking advantage of the software’s help or feedback, rather than by actively thinking through the material”

Gaming in Intelligent Tutors

- Systematic Guessing
- Help Abuse

  (cf. Wood and Wood 2000; Aleven 2001)

Gaming has also been observed in other educational settings

- Educational Games

  (Klawe 1998; Miller, Lehman, and Koedinger 1999; Magnussen and Misfeldt 2004)

- Graded-Participation Newsgroups
  (Cheng and Vassileva 2005)

- Human Teachers Giving Help
  (Arbreton 1998)

Outline

- Introduction
- Gaming and Learning
- Detecting Gaming
- Responding to Gaming
- Why Do Students Game?
- Conclusions

Does Gaming Affect Learning?

Does Gaming Affect Learning?

- We have investigated this question in five different studies, over three years
  - 300 students, ~80 minutes of usage each
  - Two suburban schools in the Northeast USA
  - 12-14 years old
  - Mathematics
  - Three lesson subjects (scatterplots, geometry, percents) with fairly different user interfaces

Gaming in Intelligent Tutors

- Methods: a combination of
  - Pre and post-tests on the tutor's subject matter
  - Quantitative Field Observations (cf. Baker et al 2004), giving an approximate proportion of time each student was gaming
    - Each student's behavior observed several times as they used tutor (by 2 observers, $\kappa = 0.83$)
    - Pre-determined order and coding categories
    - Peripheral vision
    - 20 second observation window

Relationship between Gaming and Learning

Study	N	% of students observed gaming	Coefficient	Partial r	ESS F	p	Z
Scatterplot2003	70	24%					
Scatterplot2004	110	45%					
Geometry2004	111	30%					
Scatterplot2005	34	35%					
Percents2005	41	20%					
Aggregate	**261**	**33%**	**-0.55**	**-0.16**	**n/a**	**0.03**	**-2.18**

Computed using meta-analytic techniques (I'd be happy to go into the full detail, at the end or offline)

Relationship between Gaming and Learning

Study	N	% of students observed gaming	Coefficient	Partial r	ESS F	p	Z
Scatterplot2003 Scatterplot2004 Geometry2004 Scatterplot2005 Percents2005	70 110 111 34 41	24% 45% 30% 35% 20%					
Aggregate	**261**	**33%**	**-0.55**	**-0.16**	**n/a**	**0.03**	**-2.18**

A student who gamed 1/3 of the time did 18 points worse on the post-test than a student with the same pre-test score who never gamed

Relationship between Gaming and Learning

Study	N	% of students observed gaming	Coefficient	Partial r	ESS F	p	Z
Scatterplot2003	**70**	**24%**	**-1.25**	**-0.33**	**8.07**	**0.01**	**-2.75**
Scatterplot2004	110	45%	0.20	0.05	0.26	0.61	0.51
Geometry2004	111	30%	-0.15	-0.04	0.16	0.69	-0.39
Scatterplot2005	34	35%	-0.05	-0.08	0.06	0.80	-0.25
Percents2005	**41**	**20%**	**-1.45**	**-0.34**	**5.02**	**0.03**	**-2.19**
Aggregate	261	33%	-0.55	-0.16	n/a	0.03	-2.18

Though gaming is associated with significantly worse learning

The effect is oddly unstable

I’ll discuss why a little later

Outline

- Introduction
- Gaming and Learning
- Detecting Gaming
  - Why Do We Need to Detect Gaming?
  - Detector
  - Detector Effectiveness
  - Detector Generalizability
- Responding to Gaming
- Why Do Students Game?
- Conclusions

Why do we need to detect gaming?

Why not just prevent gaming?

- Example: put in delays after every hint
- This approach was used by Carnegie Learning, and by researchers at the University of Massachusetts…
  (cf. Murray and vanLehn, 2005;
  Beck, 2005)

Why not just prevent gaming?

Problems:

- Reduces help’s effectiveness for the majority of students who don’t game
- Many students discover new ways to game (Murray and vanLehn, 2005)

Outline

- Introduction
- Gaming and Learning
- Detecting Gaming
  - Why Do We Need to Detect Gaming?
  - Detector
  - Detector Effectiveness
  - Detector Generalizability
- Responding to Gaming
- Why Do Students Game?
- Conclusions

Data

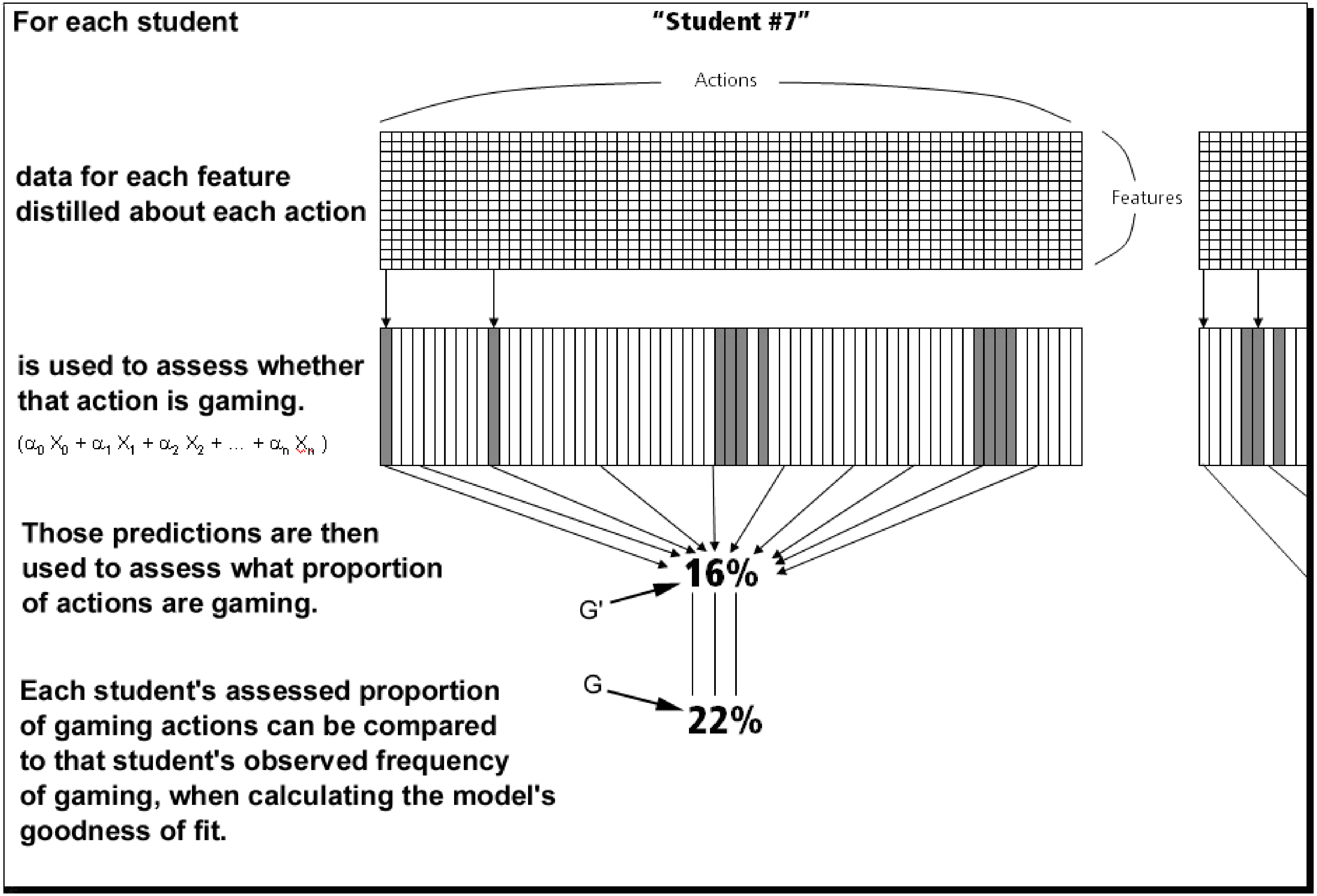
- 4 tutor lessons (scatterplots, geometry, percents, probability)
- 3 years
- 300 students in total
- 113 represented in multiple lessons
- 473 student/lesson pairs

- 128,887 actions in total

Data

- Human observations of gaming
- Pre-test, post-test
- Log file data
  - Action by action level
  - 26 features, including time taken, interface widget involved, past history on step, etc.

Modeling Framework
(Latent Response Model – e.g. Maris 1995)



Model Selection

- Specific model is selected using a combination of Fast Correlation-Based Filtering (Yu and Liu, 2003), Forward Selection, and Iterative Gradient Descent

- Searches the model space to find models which are good but not correlated with one another

Model Selection

- Models are selected and evaluated using a combination of two metrics
  - Correlation to frequency of gaming behavior
  - A' (Hanley and McNeil, 1982)

- A' is
  - the area under the ROC curve
  - the probability that if the model is given one student from each of two categories, it will accurately determine which is which
  - Chance = 0.5, Perfect = 1.0

Outline

- Introduction
- Gaming and Learning
- Detecting Gaming
  - Why Do We Need to Detect Gaming?
  - Detector
  - Detector Effectiveness
  - Detector Generalizability
- Responding to Gaming
- Why Do Students Game?
- Conclusions

Early Surprise

- In a very early attempt to build a detector
- Using only one data set (Scatterplots, 2003)
- We noticed that the detector was capturing only half the gaming students:

  The students who gamed and had poor learning ("harmful gaming")
- It wasn't capturing the students who gamed but still learned ("non-harmful gaming")

Over time, we figured out…

- We could train a detector to detect either group of gaming students
  - That detector will ignore the other group of gamers
- A detector trained on both groups (with a large data set) detects
  - most of the students who engage in “harmful” gaming
  - a moderate number of the students who engage in “non-harmful” gaming

Detector of harmful gaming

- Multiple fast errors or help requests in succession

- And also
- Gaming occurs on the steps students know least well
- Each gaming student chooses a set of steps which he/she games in every problem

- In other words, students who engage in harmful gaming appear to choose to game in order to avoid specific material they consider difficult

Detector of harmful gaming

- Ability to distinguish students who game and don’t learn, from non-gaming students
  - A’=0.84 for full data set

- Ability to distinguish students who game and don’t learn, from students who game but learn
  - A’=0.84 for full data set

Detector of non-harmful gaming

- Multiple fast errors or help requests in succession

- But also
- Answering quickly on steps the student answered slowly in the past
- Answering slowly on steps where the student has a misconception

- Non-harmful gamers appear to game time-consuming steps they already know, and work slowly and carefully on steps they don't know

Detector of non-harmful gaming

- Ability to distinguish students who game and learn, from non-gaming students
  - A'=0.81 for full data set

- Ability to distinguish students who game but learn, from students who game and don't learn
  - A'=0.82 for full data set

Outline

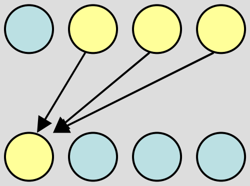
- Introduction
- Gaming and Learning
- Detecting Gaming
  - Why Do We Need to Detect Gaming?
  - Detector
  - Detector Effectiveness
  - Detector Generalizability
- Responding to Gaming
- Why Do Students Game?
- Conclusions

Detector Generalizability

- Most models and detectors of student behavior are developed within a single tutor lesson/module, or within a small-scale tutor
- Will these detectors transfer to new tutors or tutor lessons?
- If they can transfer, they'll be much more useful

Scheme

- Train on data from three lessons, test on a fourth lesson



- For all combinations of lessons (4 combinations)

Transfer lesson .vs. Training lessons

- Ability to distinguish students who game and don't learn, from non-gaming students

- Overall performance in training lessons: A' = 0.85
- Overall performance in test lessons: A' = 0.80

- Difference is NOT significant, Z=1.17, p=0.24 (using Strube's Adjusted Z)

Transfer lesson .vs. Training lessons

- Ability to distinguish students who game and don’t learn, from students who game but learn

- Overall performance in training lessons: A’ = 0.86
- Overall performance in test lessons: A’ = 0.80

- Difference is NOT significant, Z=1.37, p=0.17 (using Strube’s Adjusted Z)

Outline

- Introduction
- Gaming and Learning
- Detecting Gaming
- Responding to Gaming
  - Design of Adaptation
  - Design of Study
  - Results
- Why Do Students Game?
- Conclusions

Goals

- Adapt to when students game the system (in the harmful fashion)

- Attempting to
  - Reduce gaming
  - Improve gaming students' learning

- Minimally affecting non-gaming students' learning

Scooter the Tutor

Scooter

- Is a tutor agent who responds to gaming
  - Using graphics adapted from the Microsoft Office Assistant

- Scooter is introduced to students in a 3 minute long powerpoint-with-voiceover segment, where Scooter explains
  - who he is
  - what gaming is
  - how he will respond to gaming

Scenario

File Edit Tutor Windows Help

scenario

Samantha is trying to find out what brand of dog food her dog Champ likes best. Each day she feeds him a different brand and sees how many bowls he eats. But then her mom says that maybe her dog just eats more on days when he exercises more.

Please draw a scatterplot to show how many bowls the dog eats, given the dog's level of exercise that day.

Problem SPLOT-DB-C-0-10-0-10

OGWJMCD's skills

File Edit Tutor Windows Help

skills

- Understanding categorical variables
- Understanding numerical variables
- Determining whether a variable can be used in a bar graph
- Placing the independent variable on the X axis
- Placing the dependent variable on the Y axis
- Choosing the lower bound of an axis
- Finding the smallest value in a data set
- Finding the largest value in a data set
- Finding The range of a data set
- Choosing an appropriate scale
- Labeling the first value on the axis
- Labeling second value on the axis
- Labeling subsequent values on the axis
- Plotting the first point
- Plotting subsequent points

Worksheet

File Edit Tutor Windows Help

worksheet

Brand	Exercise (minutes)	Bowls
Bibbles and Kits	12	1
Barley's	25	0
Puppy Chew	28	3
Unleashed	5	0
Premium	18	1
Buster's	14	1
Delight	10	1
Tucker's Regular	38	3
Mad Dog	19	2
No Bones About It	27	2

Spreadsheet Calculation ON

Variable Type Tool

File Edit Tutor Windows Help

Brand

Exercise (minutes)

Bowls

Scooter the Tutor

File Edit Tutor Windows Help

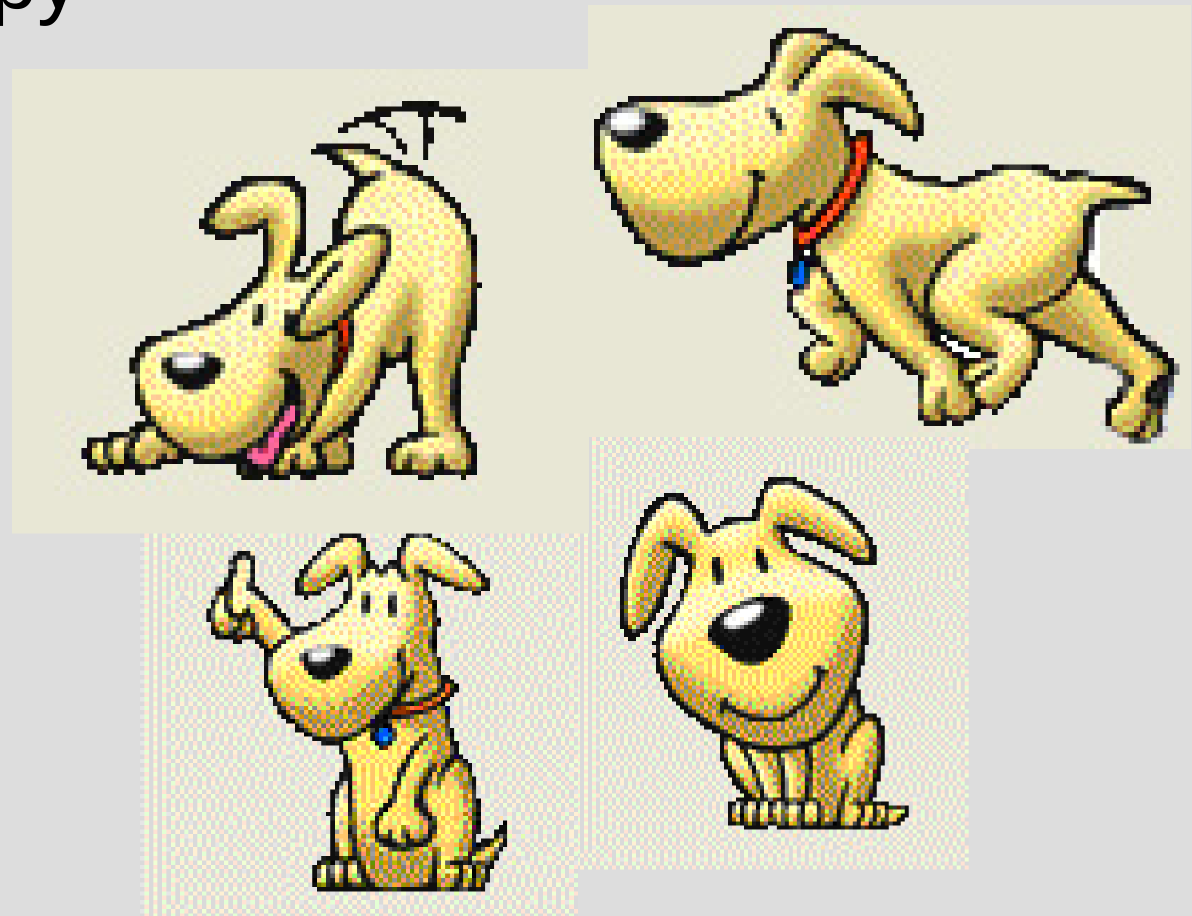
Hi! I'm Scooter the Tutor!
I'll try to help you learn how to learn!

During the Student’s Tutor Use

- Scooter responds to gaming in two ways
  - Emotional expressions
  - Supplementary exercises

Emotional Expressions

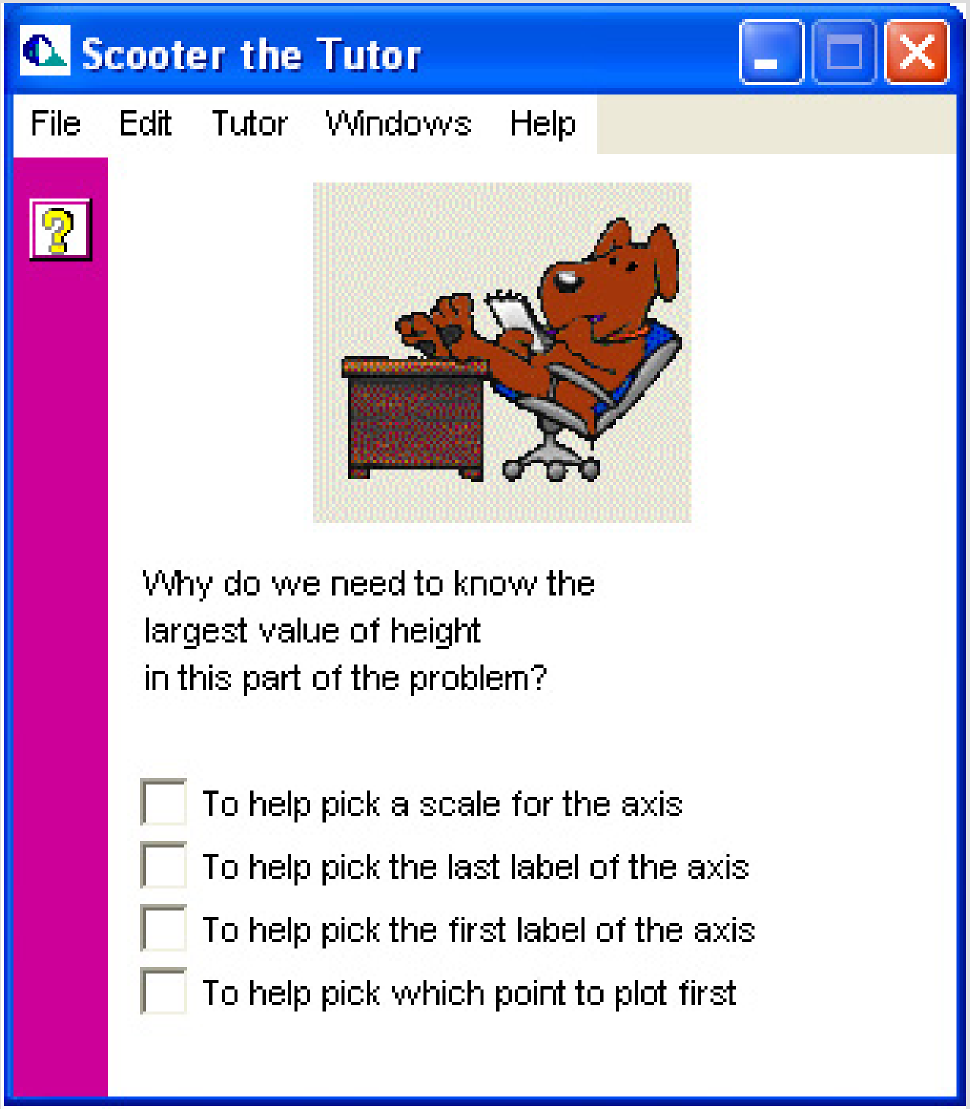
- If the student never games, Scooter looks happy



Emotional Expressions

- If the student appears to be gaming, Scooter looks increasingly displeased and becomes redder and redder (more red when more recent actions were assessed to be gaming)

Supplementary exercises



Supplementary Exercises

- Multiple Levels
- If a student gives a wrong answer, he/she receives another question

Supplementary Exercises

- 1st and 2nd levels
  - Questions which require understanding one of the concepts required to answer the step the student gamed through
  - Questions about what role the step they gamed through plays in the overall process of solving the problem
- 3rd level
  - Very easy questions, to prevent floundering

Pedagogical Agents

- Using a pedagogical agent in a tutor is not new (Grasser et al, 2003; Johnson, Rickel, and Lester, 2000; Wang et al, 2005; Manske and Conati 2005, etc etc etc etc etc)

- Evidence suggests that the mere presence of a pedagogical agent is not helpful, but that agents can affect learning positively if they enable new types of educational interactions (Graesser et al, 2003; Wang et al, 2005)

3 Hypotheses About Scooter

Hypotheses About Scooter

1. Scooter will make gaming more accountable by showing the student and their teacher who has been gaming

   Will reduce gaming overall

   (cf. Sarafino 2001)

Hypotheses About Scooter

2. Scooter's anger will invoke social norms (Reeves and Nass 1996)
   (a human tutor would become upset; some teachers *do* become upset)

   Students who receive expressions of anger will reduce their gaming after seeing the expressions of anger

Hypotheses About Scooter

3. The supplemental exercises will give students an additional opportunity to learn exactly the material they’re bypassing when they game.

   Students who receive many supplemental exercises will have better learning

Outline

- Introduction
- Gaming and Learning
- Detecting Gaming
- Responding to Gaming
  – Design of Adaptation
  – Design of Study
  – Results
- Why Do Students Game?
- Conclusions

Study Design

- Control condition
  - Regular tutor
  - 51 students (12 absent for pre or post test)

- Experimental condition
  - Tutor with Scooter
  - 51 students (17 absent for pre or post test)

Study Design

- Each student randomly assigned to one of the conditions (all students used a tutor lesson on Scatterplots)

Measures

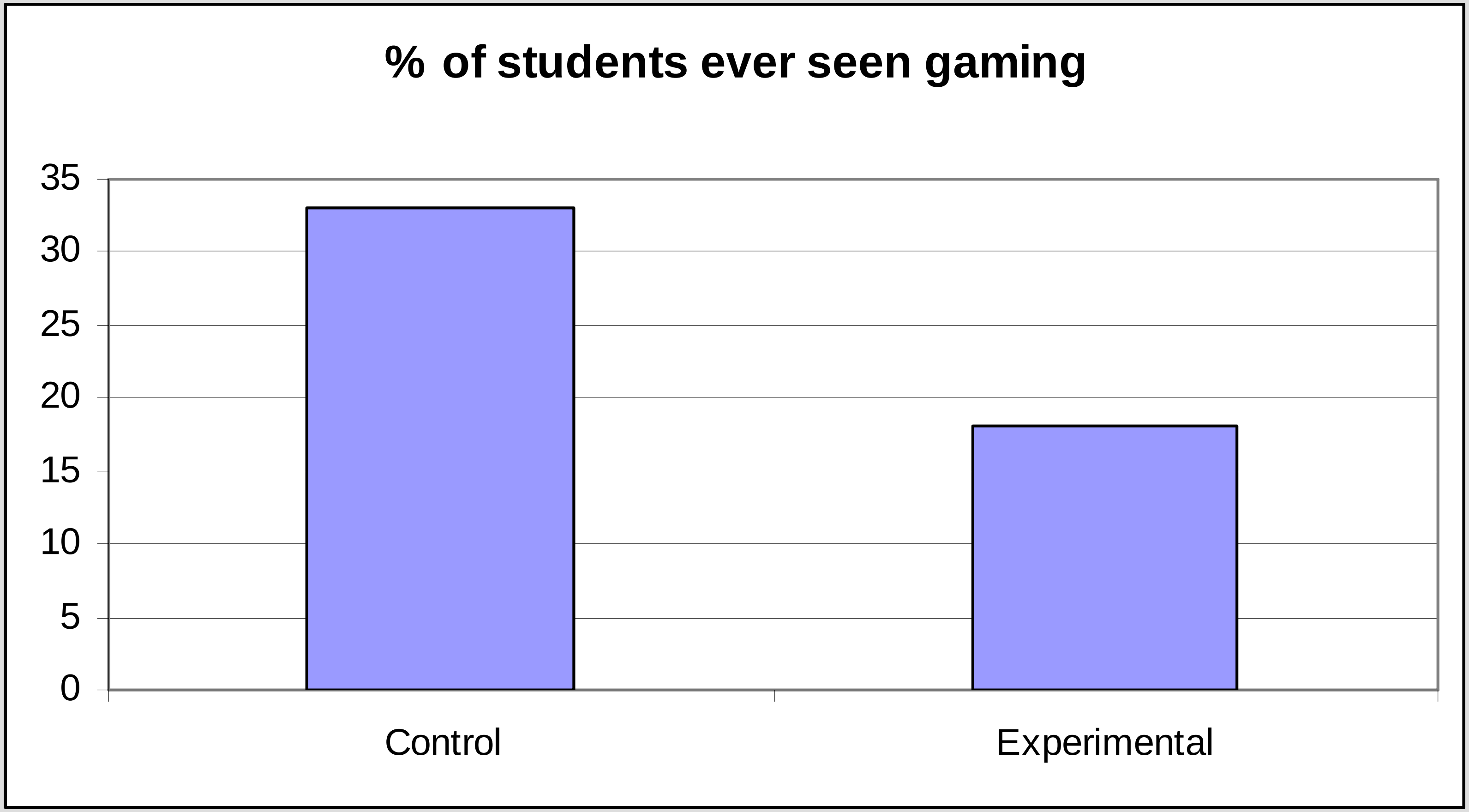
- Pre-test, post-test
  – Items counterbalanced across two tests
- Human observations of gaming
  – Using method from (Baker et al, 2004)
  – Give an estimate of the proportion of time each student spends gaming
- Log file data
  – Enables us to measure Scooter's actions
- Questionnaire Items (on opinions of Scooter)
  – I won't have time to discuss the questionnaire in detail, but I'd be happy to discuss it later

Outline

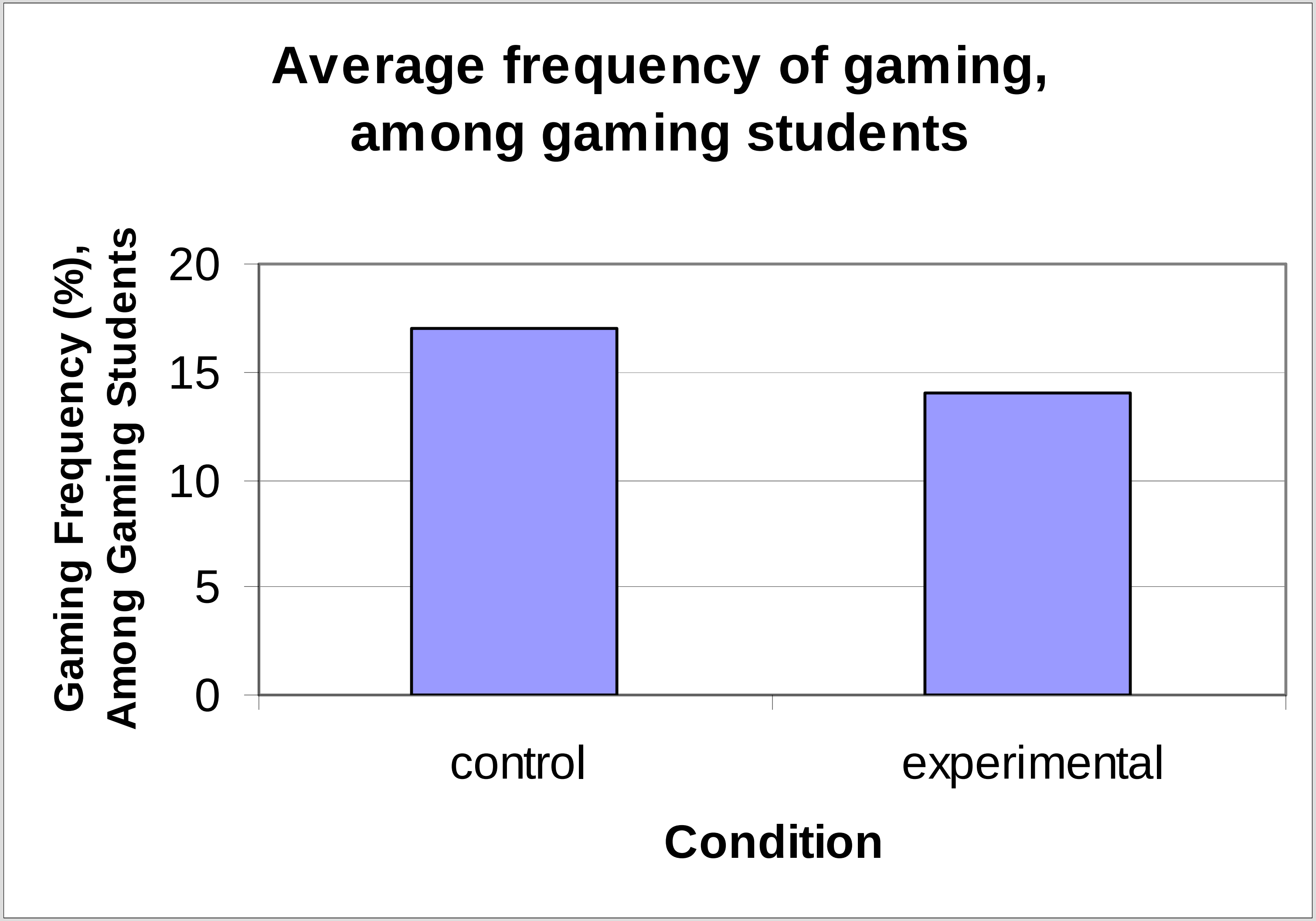
- Introduction
- Gaming and Learning
- Detecting Gaming
- Gaming and Learning (reprise)
- Responding to Gaming
  - Design of Adaptation
  - Design of Study
  - Results
- Why Do Students Game?
- Conclusions

Effects on gaming, between conditions

$\chi^2(1, N=102) = 3.30, p=0.07$



$t(23) = 0.74$, $p = 0.47$


Average frequency of gaming,
among gaming students
Gaming Frequency (%),
Among Gaming Students
20
15
10
5
0
control
experimental
Condition

Effects on learning, between conditions

Average Learning Gain Per Condition

- Control Condition: 22 point gain
- Experimental Condition: 25 point gain

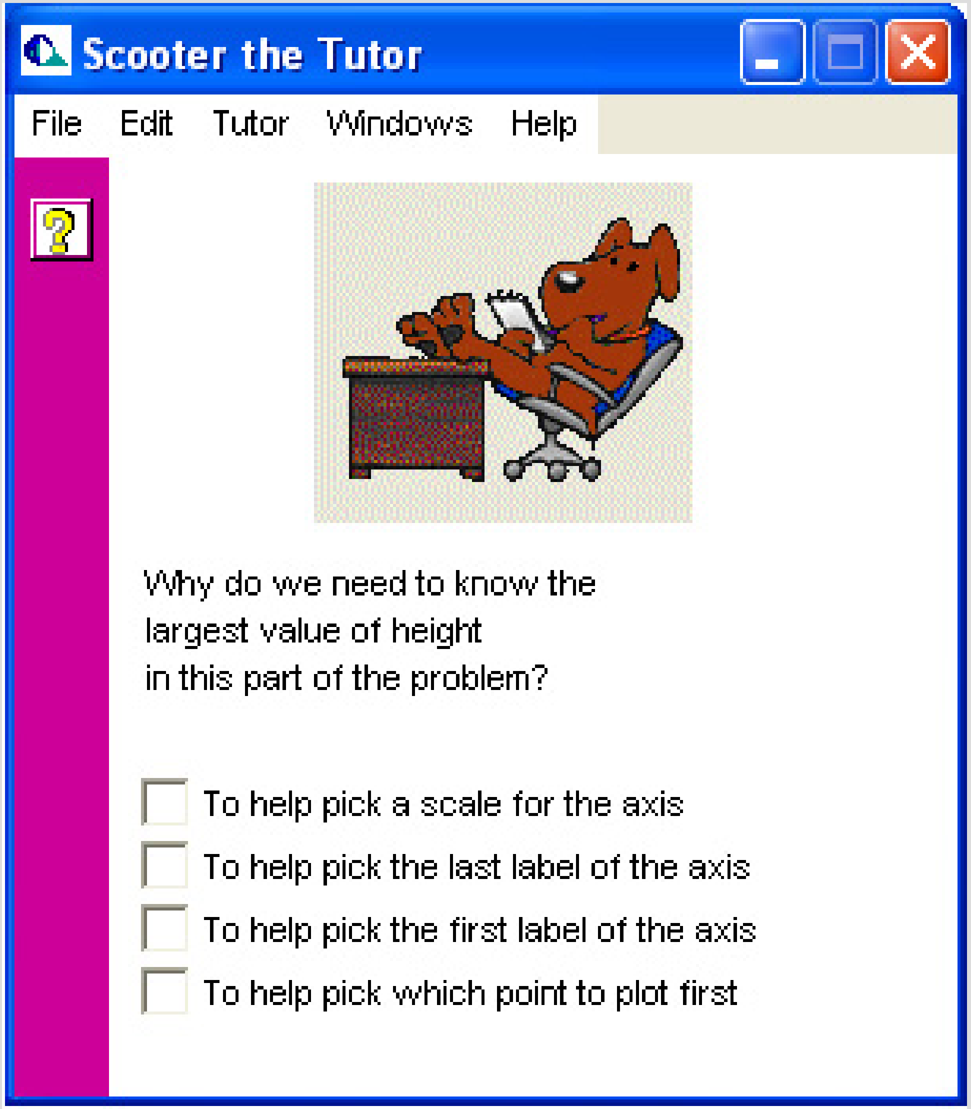
- $t(70)=0.34$, $p=0.73$

However…

- Gamers are a fairly small subset of the overall population
- Harmful gamers are an even smaller set
- The detector is not 100% -- so not all harmful gamers got interventions

- Therefore we should look at the effects of the interventions

Supplemental Exercises

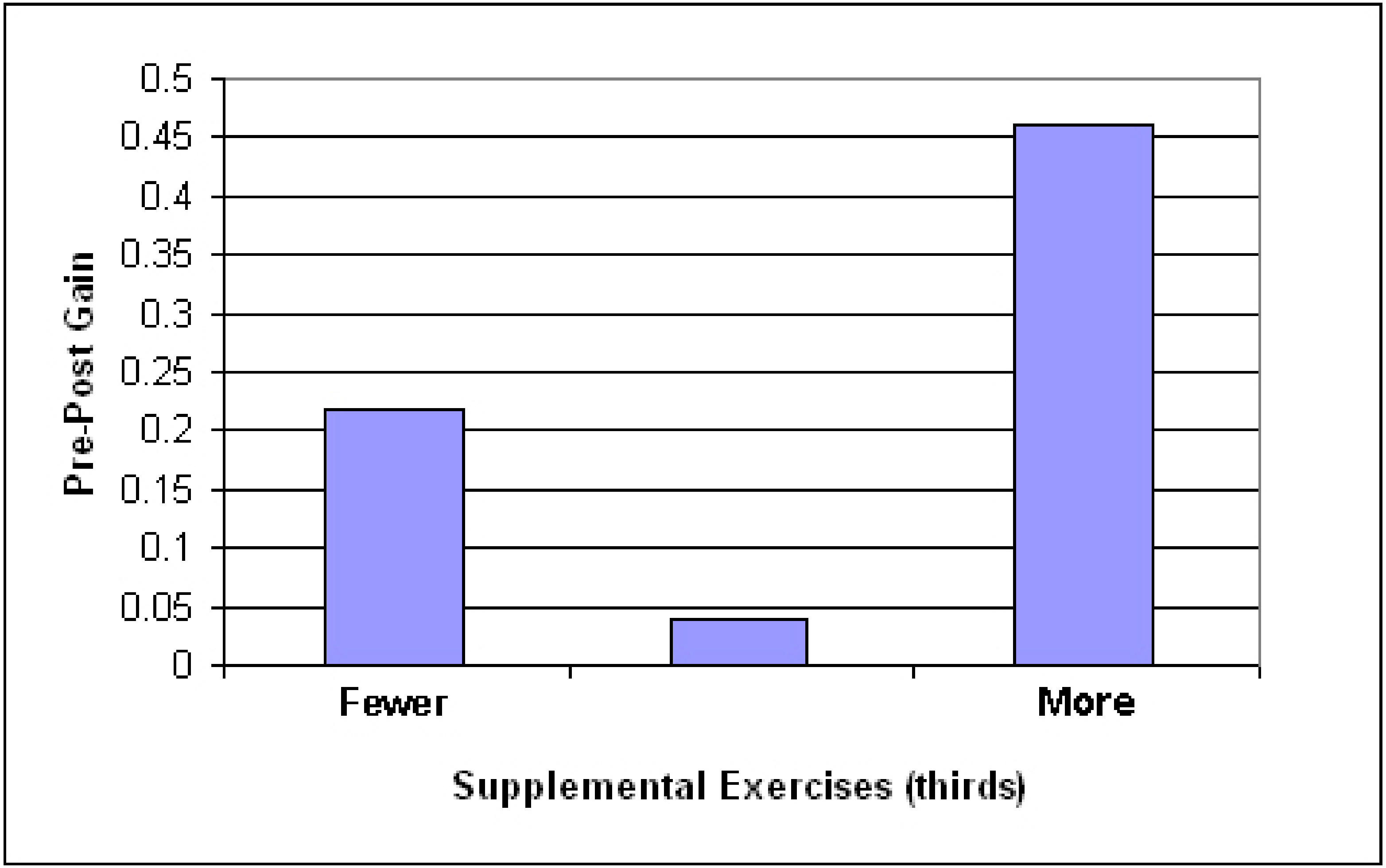


Supplemental Exercises

- Not all that many sets given
  - Max 3.2% of problem steps (12 sets)

- But targeted to exactly the steps each student is thought to be gaming on

Supplementary Exercises

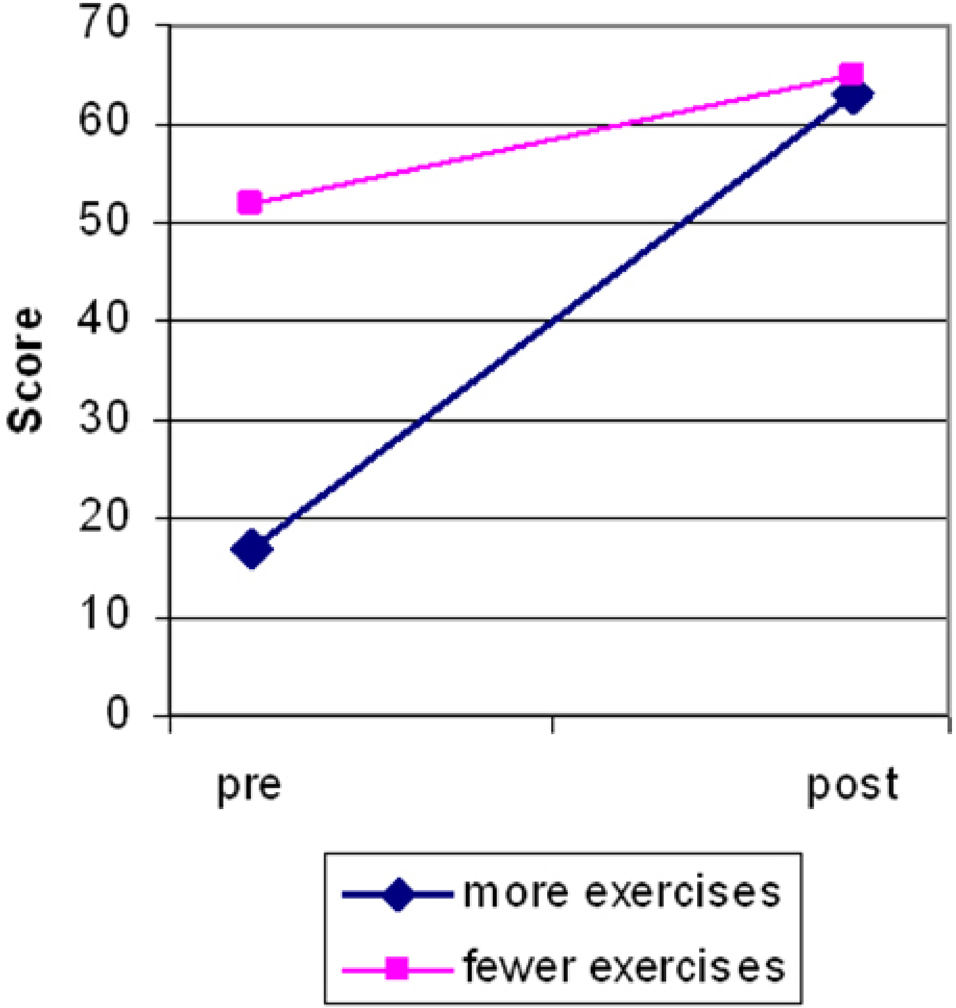


Stats

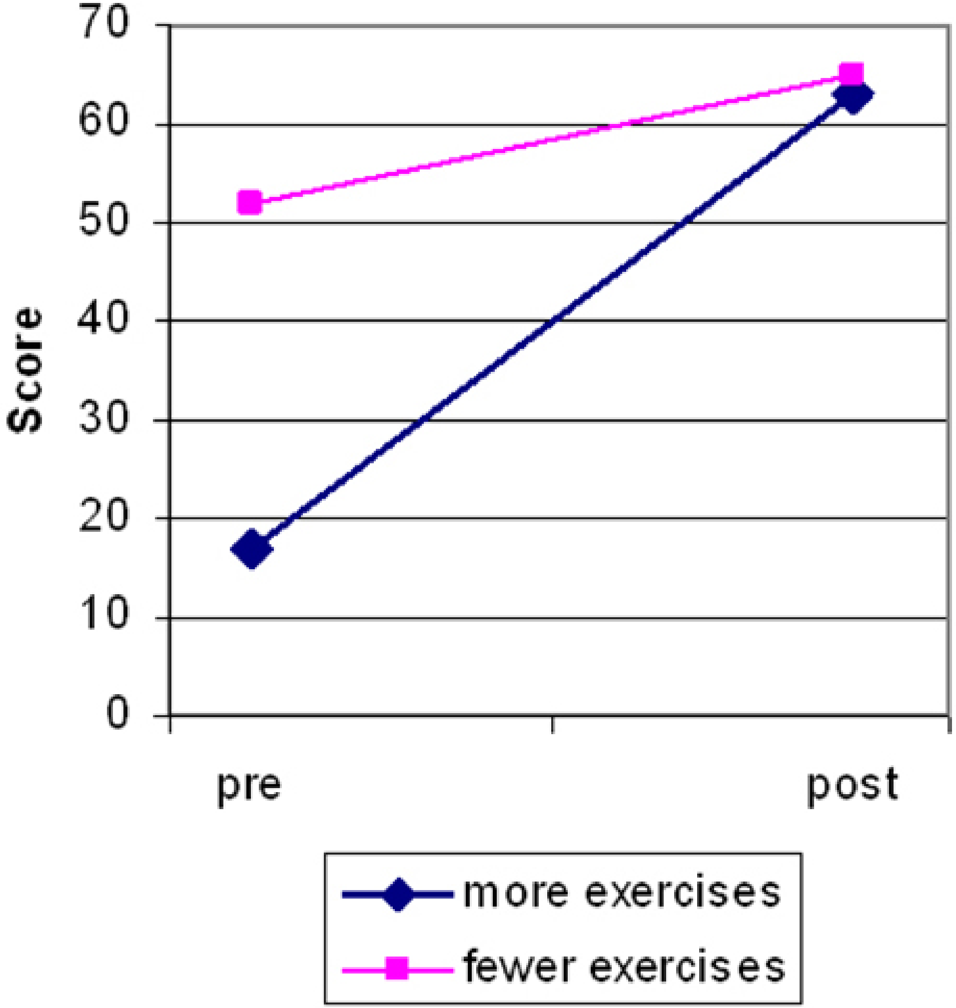
- Difference between all 3 groups
  - $F(2,36)=3.10$, $p=0.06$

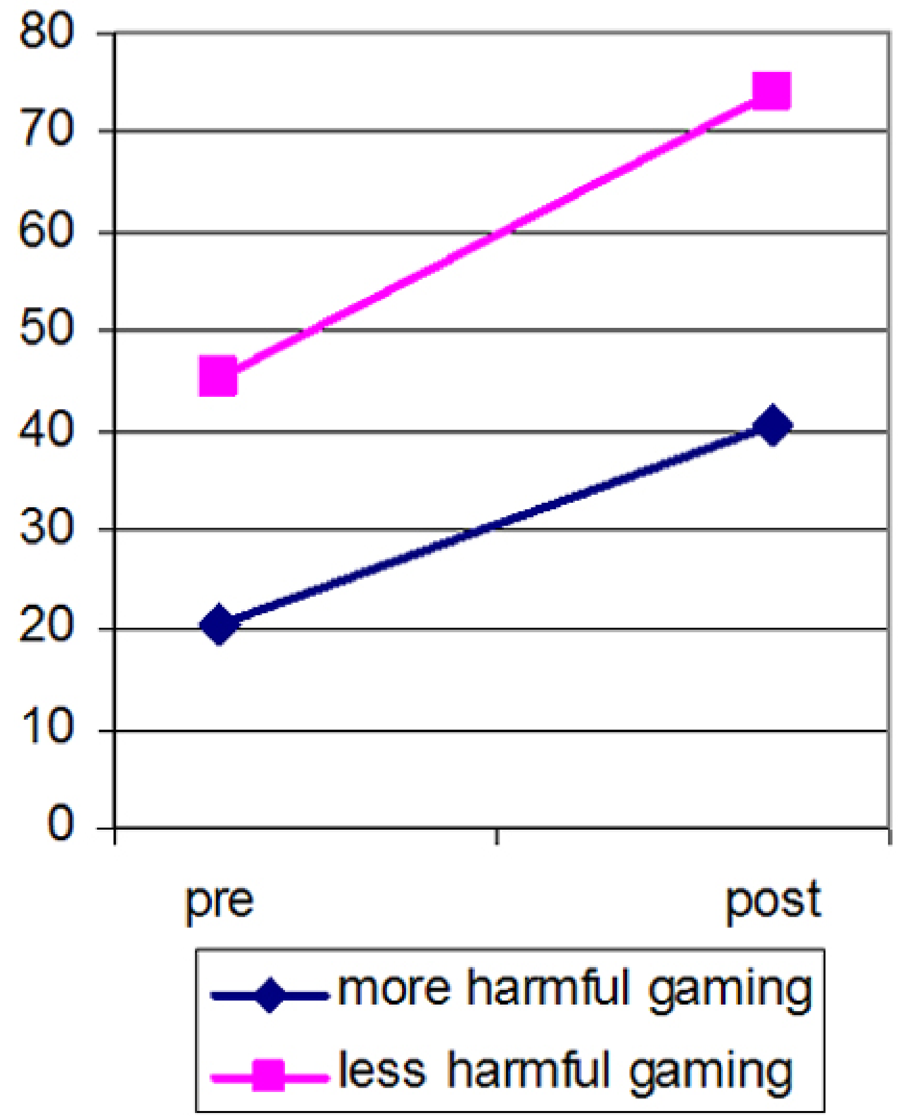
- Top third vs other two thirds
  - $t(37)=2.25$, $p=0.03$

EXPERIMENTAL



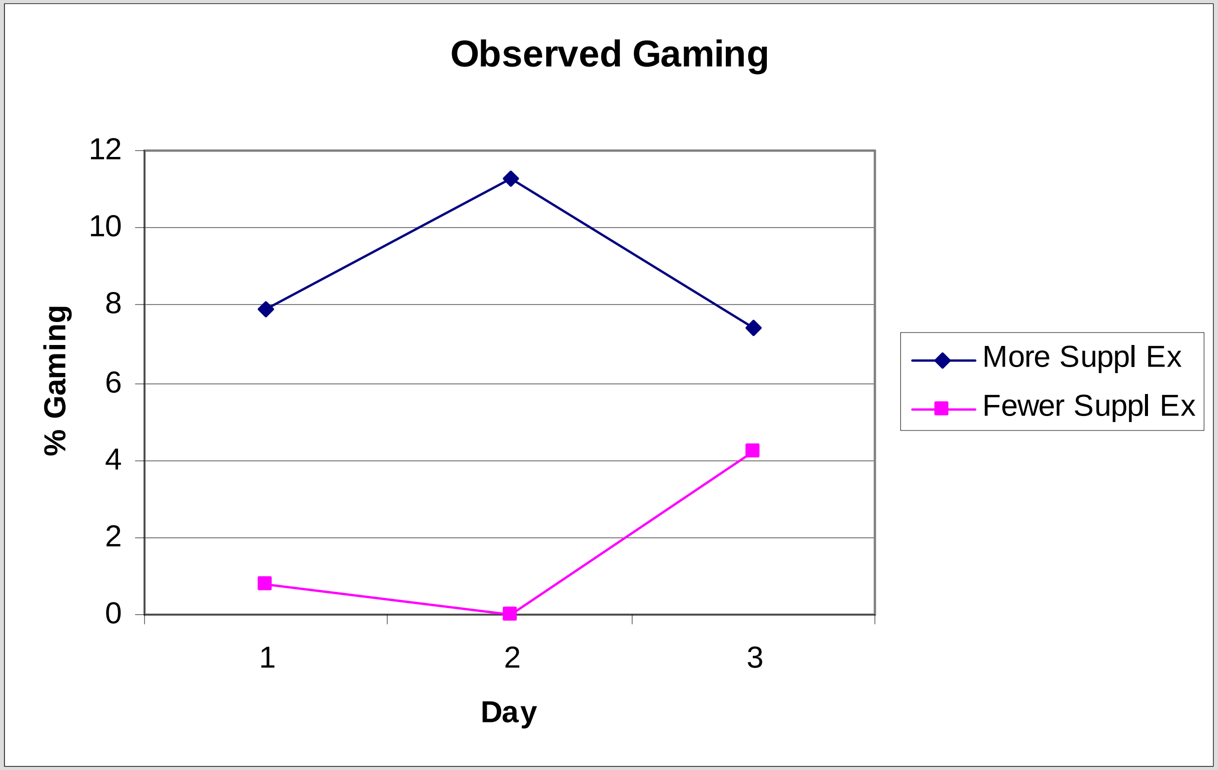
EXPERIMENTAL **CONTROL**





Interestingly…

- Students who receive more supplemental exercises do not appear to decrease their gaming over time



Expressions of Anger

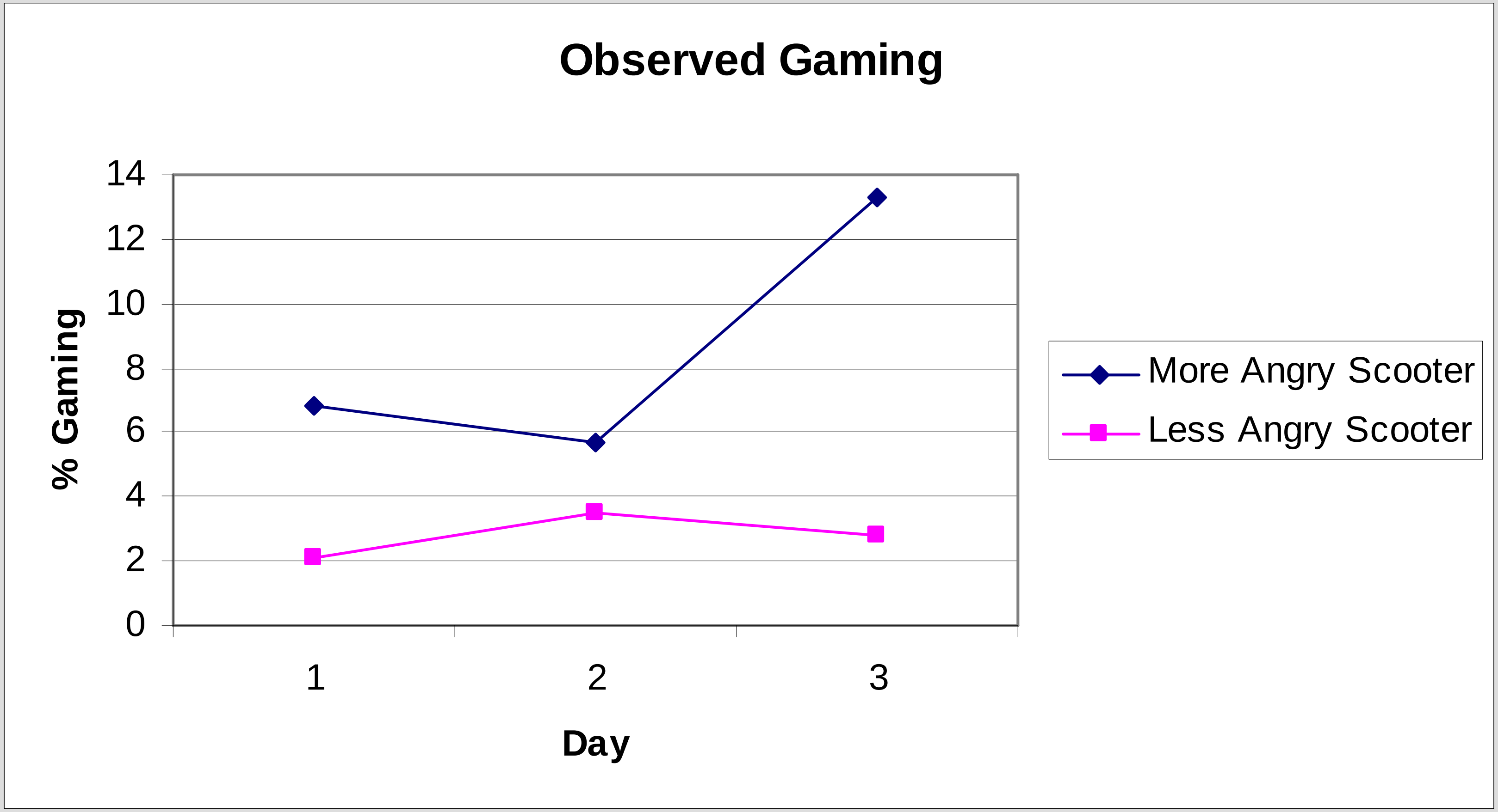
More expressions of anger were not associated with better learning

Top third .vs. rest, $t(37)=0.16$, $p=0.87$

Median split, $t(37)=0.15$, $p=0.88$

Top quartile .vs. rest $t(37)=0.48$, $p=0.63$

More expressions of anger not associated with reduction in gaming



Return To The Three Hypotheses

Hypotheses About Scooter

1. Scooter will make gaming more accountable by showing the student and their teacher who has been gaming

   Will reduce gaming

   Fewer students gamed in the experimental condition, which is consistent with this prediction, although we can't be certain that this was why gaming reduced

Hypotheses About Scooter

2. Scooter's anger will invoke social norms (Reeves and Nass 1996) (a human tutor would become angry; some teachers *do* become angry)

   Students who receive expressions of anger will reduce their gaming after seeing the expressions of anger

   Data is NOT consistent with this prediction

Hypotheses About Scooter

3. The supplemental exercises will give students an additional opportunity to learn exactly the material they're bypassing when they game.

   Students who receive many supplemental exercises will have better learning

   Data seems consistent with this prediction

Outline

- Introduction
- Gaming and Learning
- Detecting Gaming
- Responding to Gaming
  - Design of Adaptation
  - Design of Study
  - Results
- Why Do Students Game?
- Conclusions

Three studies

- Two by our group
  - With scatterplot lesson, 2004
  - With control condition of Scooter study, 2005
- One by Heffernan and Walonoski
  - With ASSISTMENTS system, 2005
- Looking at data from two systems gives us better ability to understand what quantities are associated with gaming in general, not just in a specific learning environment or population

General Method

- Use a gaming detector to get frequency of gaming for each student
- Correlate gaming to response on questionnaire items given before tutor use (13 constructs investigated)
- Differences in method at a finer-grain

Items not associated with gaming (that you might expect to be)

Not sig. associated with gaming

- Performance Goals .vs. Learning Goals
  - Everybody expected performance goals to be associated with gaming. Teachers, Researchers. Two papers (one of them ours) even put it in writing!
  - Data from each research group showed no relationship

Not sig. associated with gaming

- Anxiety

Not sig. associated with gaming

- Passive-aggressiveness
  - Interestingly, passive-aggressiveness is significantly associated with the choice to visibly go off-task (i.e. talking to a neighbor about spiderman)

Items reasonably solidly associated with gaming

- Significant relationship across studies

Associated with gaming

- Disliking Computers
- Disliking Mathematics
- Lack of Educational Self-Drive
- Frustration

- Correlations in 0.2-0.4 range

Next step

- Identify whether affect at a specific moment is related to gaming at that moment

- Ongoing work, in collaboration with Mercedes “Didith” Rodrigues

Outline

- Introduction
- Gaming and Learning
- Detecting Gaming
- Responding to Gaming
  - Design of Adaptation
  - Design of Study
  - Results
- Why Do Students Game?
- Conclusions

Summary

- My colleagues and I have
  - Determined that gaming the system leads to significantly worse learning in Cognitive Tutors
  - Determined that gaming the system splits into two categories of behavior, only one of which is associated with poorer learning
  - Studied what attitudes and motivations are associated with gaming
  - Developed a generalizable detector of gaming the system
  - Developed a software agent who responds to gaming, reducing gaming and improving learning

Current Directions

Current Directions

- Understand what student choices are associated with poorer learning in other types of interactive learning environments

- Example:
  Educational action games

Zombie Division (Hapgood 2005)

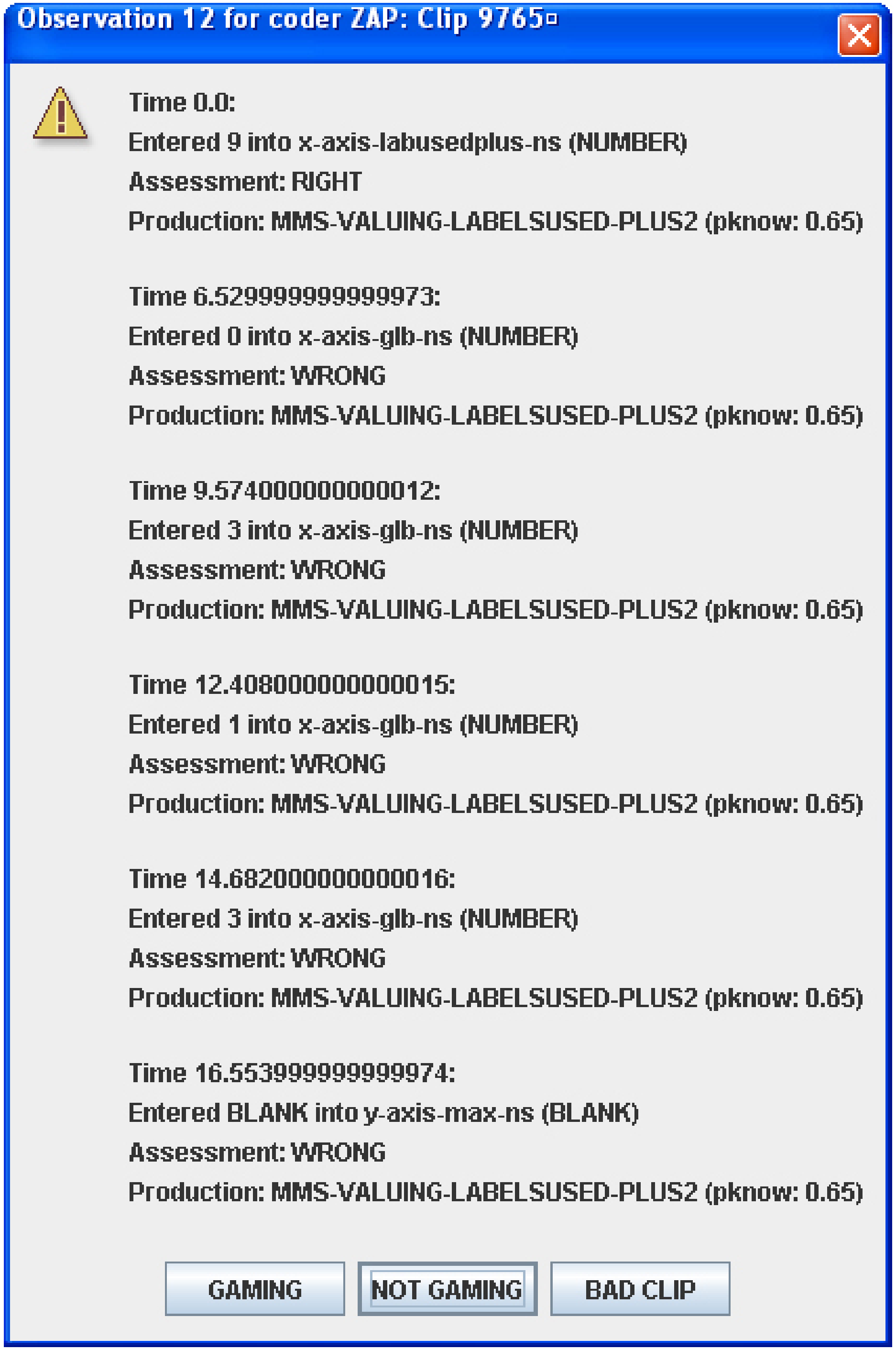
Current Directions

- Develop effective detectors of other ways students choose to use tutoring systems

- I’ve developed a system that can effectively distinguish between
  - when a student is off-task (for example, surfing the web or discussing Spiderman)
  - when a student is talking to their teacher about the subject matter
- Using only data about the student’s interactions with the software
  - i.e., no video or audio data

Current Directions

- Develop methods to rapidly generalize detectors to new systems



I've developed a technique for observing student behavior using text descriptions of sequences of behavior in low-bandwidth log files

Moderately lower interrater reliability than live observations, but 10 times faster and can be conducted retrospectively

Current Directions

- Develop methods to rapidly generalize detectors to new systems

- Currently obtaining data from SQL-Tutor as a case study for quickly building a gaming detector within another tutoring system

Current Directions

- Study what adaptations are appropriate in different types of interactive learning environments

- Dependent on first knowing what behaviors affect learning in different types of interactive learning environments

The End

- Thanks to my collaborators!

Kaska Porayska-Pomsta **Manolis Mavrikis**

Jason Walonoski **Jake Habgood**

- Thanks for your attention!

Method

- Use a gaming detector to get frequency of gaming for each student
- Correlate gaming to response on questionnaire items given before tutor use
- Differences in method at a finer-grain
  - Baker et al items adapted from existing questionnaires, motivational inventories; Walonoski and Heffernan items developed in-house
  - Baker et al used detector of harmful gaming; Walonoski and Heffernan did not find evidence of split in types of gaming, so used general gaming detector

Questionnaire Items

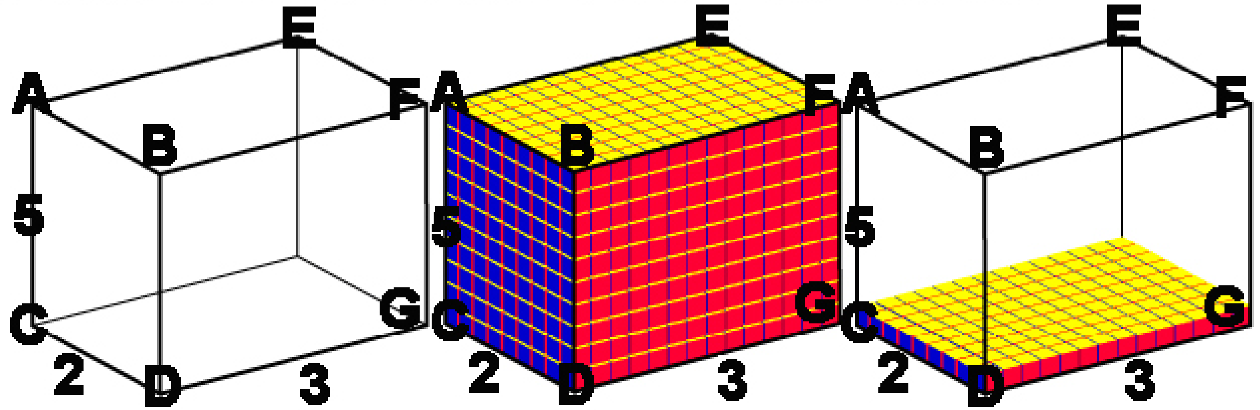
- Performance Goals .vs. Learning Goals
- Anxiety about Computers, Mathematical Task
- Attitudes towards Computers, Tutor, Math
- Believing Tutor is not Useful, Caring
- Passive-Aggressiveness
- Lack of Educational Self-Drive
- Epistemic Beliefs about Mathematics
- Frustration
- Desire for Control

Scenario

File Edit Tutor Windows Help

scenario

Use the images below to answer the following questions about the 2 by 3 by 5 rectangular prism



Problem 3-D-MEASURE-RECTANGLE-235

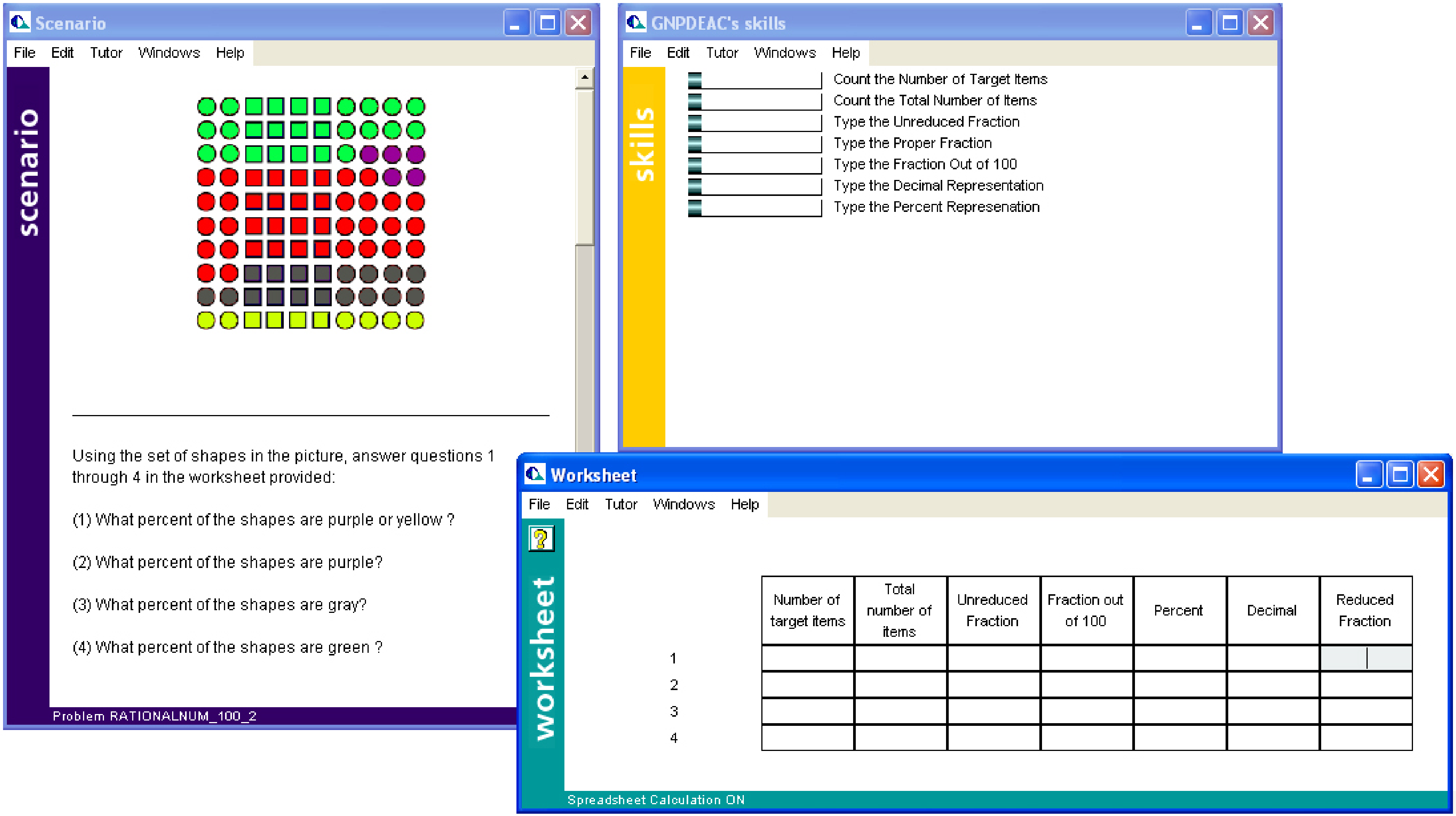
Worksheet

File Edit Tutor Windows Help

worksheet

	Name of Prism Face	Length of a Side	Length of Perpendicular Side	Shape of Prism Face	Surface Area of This Face	Number of Congruent Faces	Total Surface Area of These Faces
1							
2							
3							
4							
5							

Spreadsheet Calculation ON


Scenario
File Edit Tutor Windows Help
scenario
Using the set of shapes in the picture, answer questions 1 through 4 in the worksheet provided:
(1) What percent of the shapes are purple or yellow ?
(2) What percent of the shapes are purple?
(3) What percent of the shapes are gray?
(4) What percent of the shapes are green ?
Problem RATIONALNUM_100_2
GNPDEAC's skills
File Edit Tutor Windows Help
skills
Count the Number of Target Items
Count the Total Number of Items
Type the Unreduced Fraction
Type the Proper Fraction
Type the Fraction Out of 100
Type the Decimal Representation
Type the Percent Represenation
Worksheet
File Edit Tutor Windows Help
worksheet
Number of target items
Total number of items
Unreduced Fraction
Fraction out of 100
Percent
Decimal
Reduced Fraction
1
2
3
4
Spreadsheet Calculation ON

Scenario

File Edit Tutor Windows Help

scenario

Each row of veggies represents a set of veggies in a box. For each question, pretend that you close your eyes and reach into the box for that question. Please answer the question that goes along with each box.

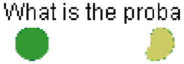
(1) What is the probability of picking a pea?

(2) What is the probability of picking a kernel of corn?

(3) What is the probability of picking a limabean?

(4) What is the probability of picking a carrot?

(5) What is the probability of picking a plum?



(6) What is the probability of picking a limabean?

(7) What is the probability of picking a pea?

Problem PROBABILITY-FRUIT-1

Worksheet

File Edit Tutor Windows Help

worksheet

	Number of target veggies	Total number of veggies	Unreduced Probability	Greatest common factor	Reduced Probability
1					
2					
3					
4					
5					
6					
7					

Spreadsheet Calculation ON

Lessons and Quantity of Data

Lesson	Number of students	Number of actions in logs
SCATTERPLOT	268 (2003-2005)	71,236
GEOMETRY	111 (2004)	30,696
PERCENTS	53 (2005)	16,196
PROBABILITY **(no pre-test data available)**	41 (2004)	10,759

Log File Data

- For each student action, we distilled 26 features, including:
  - Info about the specific action
    - Correct, wrong, help request?
    - Interface Widget Type
  - Info about how the student's action compared to similar actions by other students
    - Within this problem step, how much faster or slower than average was the action?
  - Info about the historical context of the action, for this student
    - Such as the student's history of errors, successes, and help use on this skill

Model Selection

- Specific model is selected using a combination of Fast Correlation-Based Filtering (Yu and Liu, 2003), Forward Selection, and Iterative Gradient Descent

- FCBF selects a set of single-parameter models which are each good but not correlated to one another
- Forward Selection used to iteratively add parameters to each model
- Parameter values are selected using iterative gradient descent
- Final model size in some cases capped at 6 parameters, in other cases determined using leave out one cross-validation

Which Measure of Gaming?

- Human Observations
  - Conflates two types of gaming
- Detector Observations
  - Are being used to drive interventions, creating serious risk of bias in analyses

- In these analyses, I use human observations as the measure of gaming, to avoid bias

Relationship between Gaming and Learning

Study	N	% of students observed gaming	Coefficient	Partial r	ESS F	p	Z
Scatterplot2003	70	24%	-1.25	-0.33	8.07	0.01	-2.75
Scatterplot2004	110	45%	0.20	0.05	0.26	0.61	0.51
Geometry2004	111	30%	-0.15	-0.04	0.16	0.69	-0.39
Scatterplot2005	34	35%	-0.05	-0.08	0.06	0.80	-0.25
Percents2005	41	20%	-1.45	-0.34	5.02	0.03	-2.19
Aggregate	**261**	**33%**	**-0.55**	**-0.16**	**n/a**	**0.03**	**-2.18**

Coefficient for each study was weighted by sample-size and averaged

Partial correlation for each study was converted to fisher Z, weighted by sample-size, averaged, and re-converted to correlation

Strube's Adjusted Z

Detector of harmful gaming

	param 1	param 2	value	description
GH1	howmanywrong	wrongpct	0.08	GH: Many errors across problems
GH2	pknow if first action, -1 otherwise	wrongpct	1.25	GH: History of many errors and yet a high probability the student knows the skill (ie lots of errors on some problems, other times correct on the first try)
GH3	point	wrongpct	-2.22	Not GH: Lots of errors while plotting points
GH4	pknow	recent8help	0.66	GH: Asking for a lot of help, and then reaching a step which the system knows the student knows
GH5	notfirstaction	timelast3SD	-0.72	GH: Very fast actions after making at least one error
GH6	timelast3SD	timelast3SD	-0.34	Not GH: Very fast answers or very slow answers
GH7	timelast3SD	wrongpct	0.37	Not GH: Very fast answers on steps with a high frequency of errors across problems

Why is meta-analysis necessary?

- In analyzing data from four lessons
- It is not appropriate to simply collapse data from different lessons together
  - Would bias in favor of detectors that perform better on lessons with more data
  - Would underestimate A', because gaming occurs with different average frequency in different lessons

Approach

- Determine A’ and correlation for each detector for each test lesson separately (using standard formulas)
- Compare detectors to each other within each lesson
- Aggregate data across lessons to conduct overall statistical significance tests

Approach

- Compare detectors to each other within each lesson
  - Two A' values can be compared to one another using a Z test with Hanley and McNeil's (1982) method for calculating standard error of A', producing a Z score

Approach

- Aggregate data across lessons to conduct overall statistical significance tests
  - Using Strube's (1985) Adjusted Z method

- Avoids overemphasizing information from students who are represented in multiple data sets

- Explicitly accounts for intercorrelation between data sets with overlapping participants

Lessons detector trained on	A' (GAMED-HURT vs NON-GAMING) when detector tested on lesson.			
	Scatterplot	Percents	Geometry	Probability
Percents, Geometry, Probability	**0.67**	0.91	0.77	0.96
Scatterplot, Geometry, Probability	0.75	**0.86**	0.76	0.99
Scatterplot, Percents, Probability	0.81	0.93	**0.69**	0.92
Scatterplot, Percents, Geometry	0.75	0.92	0.77	**0.99**

Lessons detector trained on	A' (GAMED-HURT vs GAMED-NOT-HURT) when detector tested on lesson.			
	Scatterplot	Percents	Geometry	Probability
Percents, Geometry, Probability	**0.6**	0.8	0.92	0.99
Scatterplot, Geometry, Probability	0.69	**0.75**	0.94	0.99
Scatterplot, Percents, Probability	0.74	0.9	**0.84**	0.99
Scatterplot, Percents, Geometry	0.68	0.8	0.89	**0.99**

Relationship between Harmful Gaming and Learning

Study	N	Partial r	ESS F	p	Z
Scatterplot2003	70	-0.17	2.03	0.16	-1.41
Scatterplot2004	110	-0.11	0.63	0.43	-0.79
Geometry2004	**111**	**-0.18**	**3.58**	**0.06**	**-1.86**
Scatterplot2005	**34**	**-0.42**	**6.99**	**0.01**	**-2.59**
Percents2005	**41**	**-0.66**	**27.49**	**<0.01**	**-4.78**
Aggregate	**261**	**-0.29**	**n/a**	**<0.01**	**-5.02**

Relationship between Non-Harmful Gaming and Learning

Study	N	Partial r	ESS F	p	Z
Scatterplot2003	70	-0.01	0.003	0.96	-0.06
Scatterplot2004	**110**	**0.23**	**6.30**	**0.01**	**2.44**
Geometry2004	**111**	**0.12**	**1.62**	**0.21**	**2.11**
Scatterplot2005	**34**	**0.35**	**4.62**	**0.04**	**1.26**
Percents2005	**41**	**-0.44**	**8.42**	**0.01**	**-2.81**
Aggregate	261	0.05	n/a	0.18	1.32

Group	Learning Gain
Experimental condition: more supplementary exercises	46 points
Control condition: more harmful gaming 2004: more harmful gaming (Baker et al 2005) 2003: more harmful gaming (Baker 2005) Average without supplementary exercises	20 points 18 points 25 points 22 points

$t(47)=2.09, p=0.04$

Another Opportunity

- It might be helpful to identify which steps non-gaming students are specifically floundering on, and give supplementary exercises on those steps

Another Opportunity

- It might be helpful to identify which steps non-gaming students are specifically floundering on, and give supplementary exercises on those steps

- To be explored in a follow-up study near you!

Why do students game/go off-task?

	Performance Goals	Anxiety About Using Computers	Anxiety About Using the Tutor	Lying/ Answering Carelessly	Disliking Computers	Disliking the Tutor
Off-Task Behavior	0.11	0.13	0.04	-0.03	**0.22**	0.12
Gaming the System (harmful fashion)	0.00	-0.02	-0.04	0.06	**0.19**	**0.20**

Why do students game/go off-task?

	Belief that Computers / the Tutor are not useful	Belief that Computers/ the Tutor are uncaring	Tendency towards passive-aggressiveness	Belief that Computers/ the Tutor reduce control	The student is not self-driven	Disliking math
Off-Task Behavior	0.02	-0.03	**0.20**	0.00	*0.17*	**0.27**
Gaming the System (harmful fashion)	0.16	0.13	0.10	0.04	**0.25**	**0.21**

Items

- Whether the student had performance goals or learning goals

  (Example: “We are considering adding a new feature to the computer tutors, to give you more control over the problems the tutor gives you. If you had your choice, what kind of problems would you like best?

  A) Problems that aren’t too hard, so I don’t get many wrong.

  B) Problems that are pretty easy, so I’ll do well.

  C) Problems that I’m pretty good at, so I can show that I’m smart.

  D) Problems that I’ll learn a lot from, even if I won’t look so smart.”)

- The student’s level of anxiety about using the tutor

  (Example: “When you are working problems in the tutor, do you feel that other students understand the tutor better than you?”)

Items

- The student’s level of anxiety about using computers (Example: “When you use computers in general, do you feel afraid that you will do something wrong?”)
- How much the student liked using the tutor (Example: “How much fun were the math problems in the last computer tutor lesson you used?”)
- The student’s attitude towards computers (Example: “How much do you like using computers, in general?”)
- If the student was lying or answering carelessly on the questionnaire

  (Example: “Is the following statement true about YOU? ‘I never worry what other people think about me.’ TRUE/FALSE”)

Items

- If the student believes that computers in general, and the tutor in specific, are not very useful.

  (Example: “Most things that a computer can be used for, I can do just as well myself.”)

- If the student believes that computers/the tutor don’t/can’t really care how much he/she learns.

  (Example: “I feel that the tutor, in its own unique way, is genuinely concerned about my learning.”)

- If the student has a tendency towards passive-aggressiveness

  (Example: “At times I tend to work slowly or do a bad job on tasks I don’t want to do”)

Items

- If the student believes that computers/the tutor reduce his/her sense of being in control

  (Example: “Using the tutor gives me greater control over my work”)
- If the student is not educationally self-driven

  (Example: “I study by myself without anyone forcing me to study.”)
- If the student dislikes math

  (Example: “Math is boring”)

Long-Term Goal

Long-Term Goal

- A general framework for Learner-Computer Interaction
  - How do students choose to use Interactive Learning Environments?
    - How do these choices differ between environments?
    - Which student choices affect learning?
  - How can ILEs automatically detect the differences in student choices?
  - How should different types of ILEs respond to student choices?

Which might look something like

- A model
- You give it
  - a detailed semantic description of an ILE's interface
  - a bunch of log data from students using that ILE
- It gives you back
  - a list of student behaviors that are negatively affecting learning within your system
  - generalizable detectors of those behaviors
  - useful suggestions for how your system could be re-designed to adapt to those behaviors