Visual Data Mining as a Tool for Educational Data

Jürgen Symanzik

Utah State University, Logan, Utah, USA



e-mail: symanzik@math.usu.edu

http://www.math.usu.edu/~symanzik

Contents

- Terms, Citations, and Definitions
- Case Study 1: The “Refraction Game”
- Case Study 2: An iPad Study
- Conclusion

Terms

- Interactive & Dynamic Statistical Graphics (DSG)
- Exploratory Data Analysis (EDA)
- Exploratory Spatial Data Analysis (ESDA)
- Visual Data Mining (VDM)
- Visual Analysis/Visual Analytics (VA)
- Data Mining (DM)

Citations

- John W. Tukey (1977):

  EDA *"is detective work - numerical detective work - or counting detective work - or graphical detective work."*

- Edward J. Wegman (2000):

  "Data Mining is exploratory data analysis with little or no human interaction using computationally feasible techniques, i.e., the attempt to find interesting structure unknown a priori."

Visual Data Mining (1)

- Working Definition for VDM:
  - Find structure (cluster, unusual observations) in large and not necessarily homogeneous data sets based on human perception using graphical methods and user interaction
  - Goal or expected outcome of exploration usually unknown in advance

Visual Data Mining (2)

- First uses of the term VDM:
  - Cox, Eick, Wills, Brachman (1997): Visual Data Mining: Recognizing Telephone Calling Fraud, *Data Mining and Knowledge Discovery*, 1:225-231.
  - Inselberg (1998): Visual Data Mining with Parallel Coordinates, *Computational Statistics*, 13(1):47-63.

Visual Data Mining Concepts

- Use existing visualization techniques, such as
  - Scatterplots and Scatterplot Matrices
  - Parallel Coordinate Plots
  - Heatmaps
  - Mosaic Plots
  - Brushing and Linked Brushing/Linked Views
  - Rotations and Projections
  - Grand Tour
  - “Small Multiples”, …
- Develop customized visualization techniques

Main References

• Symanzik, J. (2012): Interactive and Dynamic Graphics [Revised], In: Gentle, J. E., Haerdle, W. K., Mori, Y. (Eds.), Handbook of Computational Statistics --- Concepts and Methods, Vol. 1 (Second Revised and Updated Edition), Springer, Berlin/Heidelberg, 335-373.

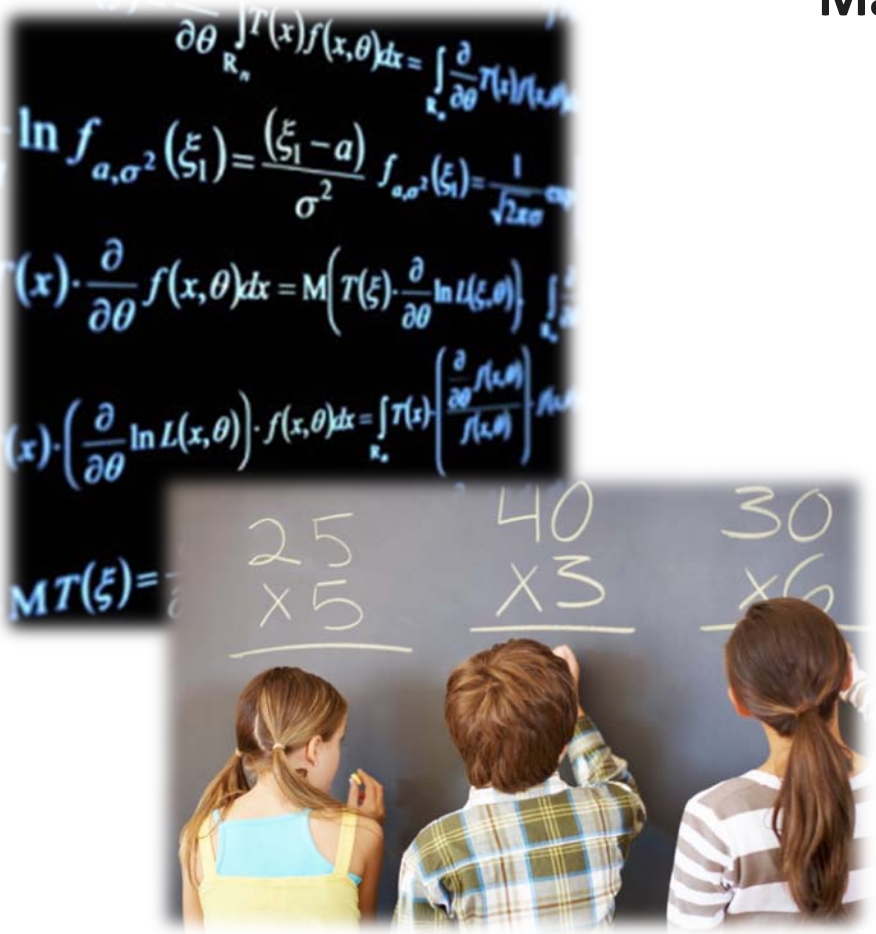
• Symanzik, J. (2011): Interactive and Dynamic Statistical Graphics, In: Lovric, M. (Ed.), International Encyclopedia of Statistical Science, Springer, Berlin/Heidelberg, 674-679.

Case Study1: The “Refraction Game”

Published as:

Aghababyan, A., Symanzik, J., Martin, T. (2013): Visualization of “States” in Online Educational Games, Proceedings of the 59th World Statistics Congress of the International Statistical Institute, 2013, International Statistical Institute (ISI), The Hague, The Netherlands.

Background: The Need



Mathematics Education

- Mathematics proficiency is critical for success in school
- Algebra is a gateway to high school & college
- Rational numbers are core concept in elementary school mathematics
- Fractions are one of the most challenging areas of mathematics

Educational Games

Electronic Games:

- Routine part of childhood

Educational Games:

- Rising popularity

Design:

- Unstructured environments
- Educational concepts
- Trial and error
- Fun interfaces

Justification:

- Students love games
- Naturally motivated

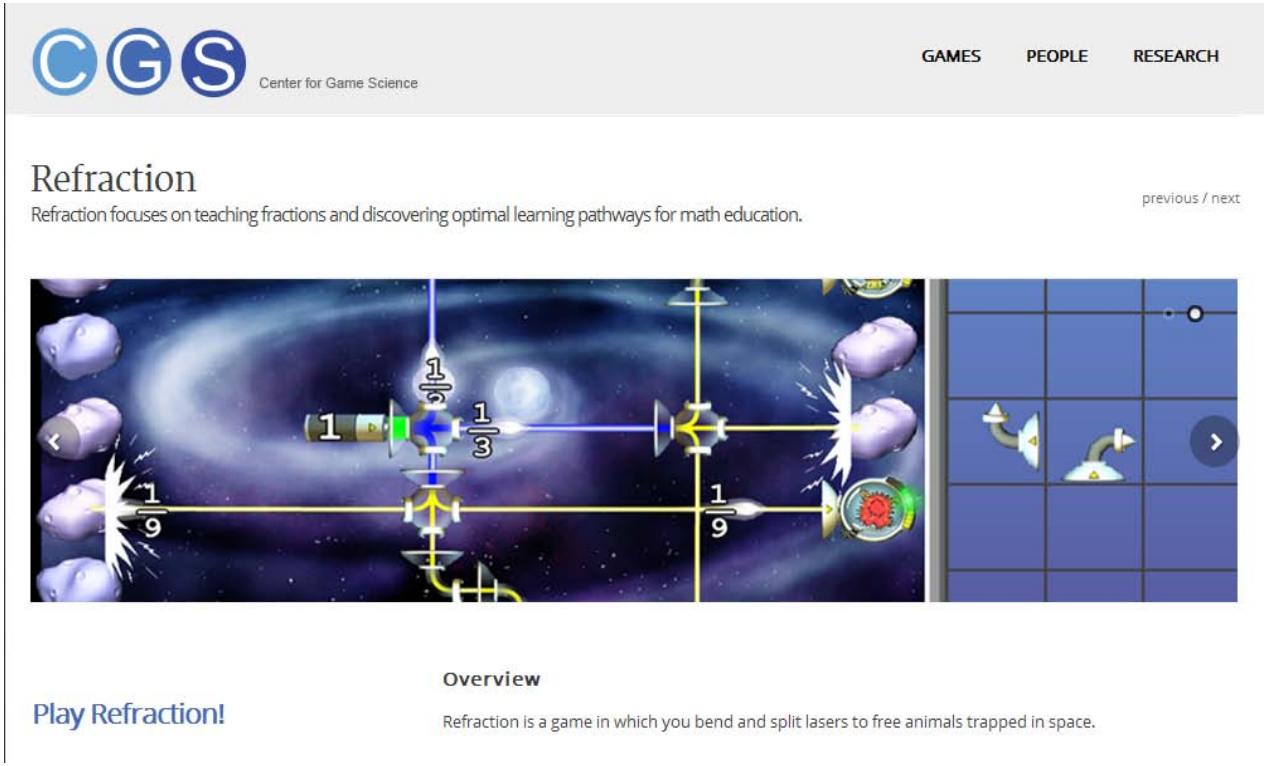






What is Refraction?

- Free, online educational game, developed at the Center for Game Science (CGS), University of Washington
- Aimed at developing pre-algebra concepts in elementary and middle school students



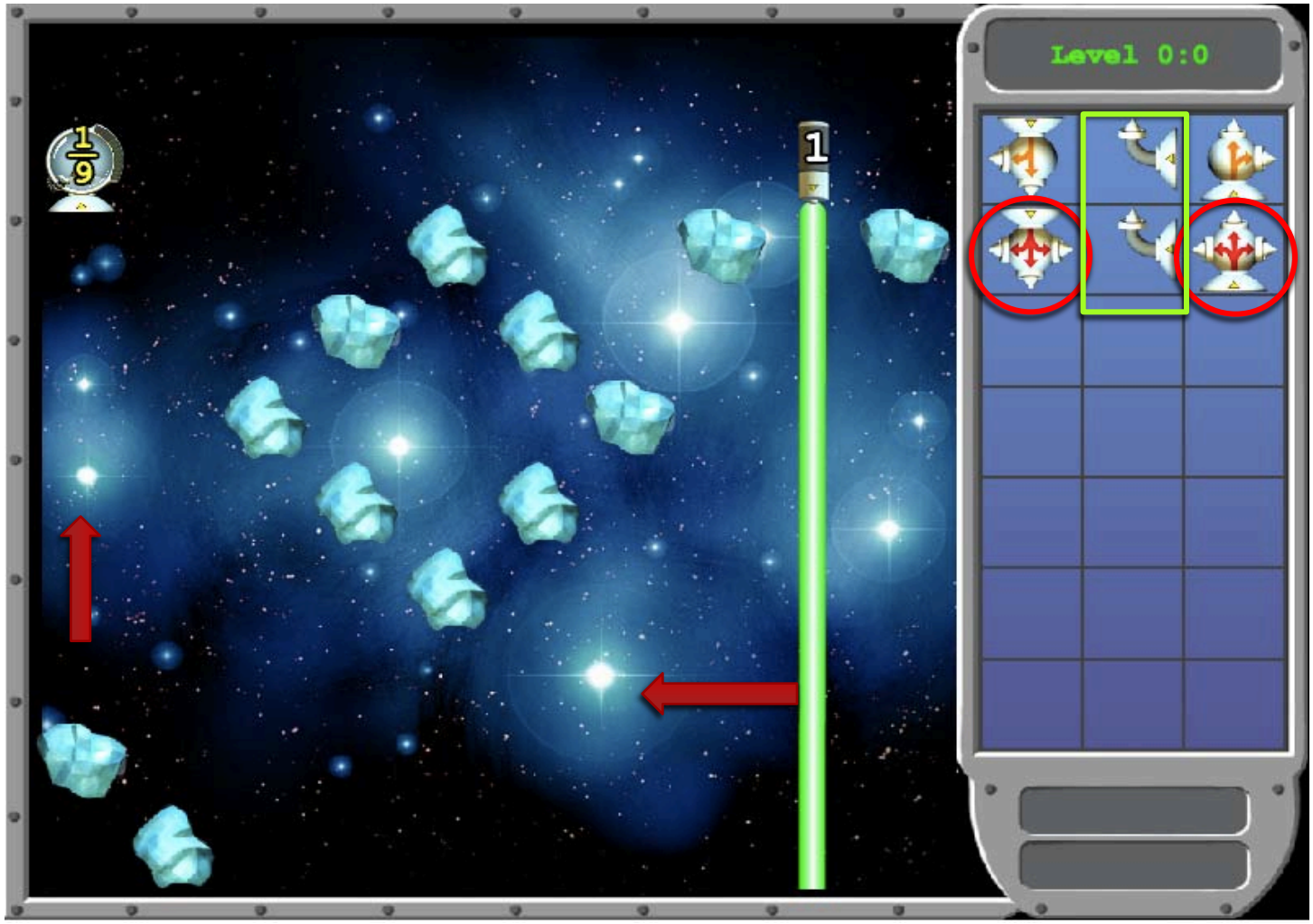
http://centerforgamescience.org/portfolio/refraction/

- **Research Questions**
  - Discovering optimal pathways for learning early mathematics
  - Detecting what it looks like when children learn fractions
  - Categorization of ways students learn by playing



Refraction Level 189

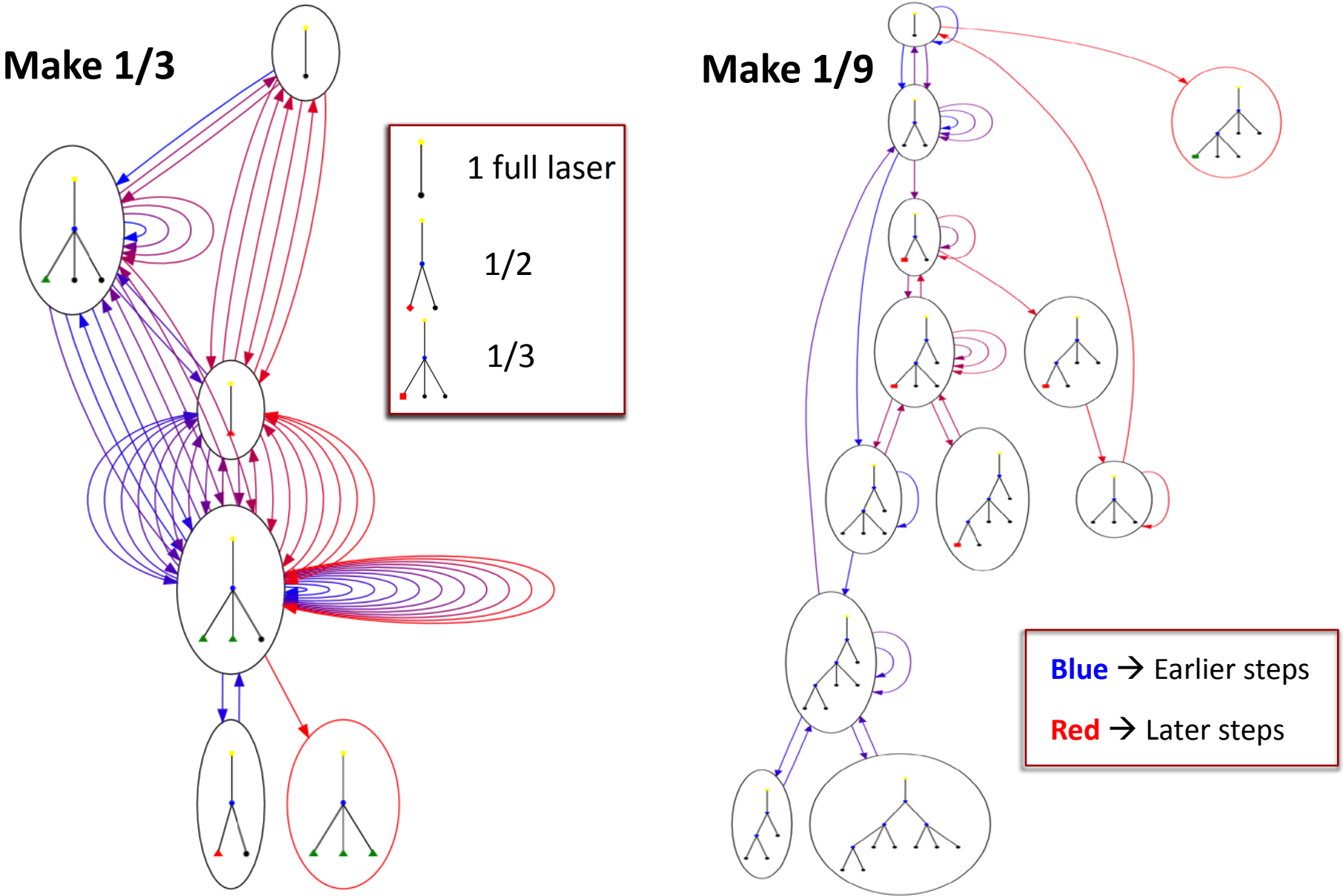
http://centerforgamescience.org/portfolio/refraction/



Fraction Man Visualization

- Motivation: Display gameplay of a student
- “States” & “transitions” of the game shown as directed graph
- State: represented as vertex (ellipse)
- Transition: represented as edge
- Use specific symbols for the splitters
- Addition of benders does not result in a new state
- Sequential blue-red color scheme used to display order of transitions

Fraction Man Visualization



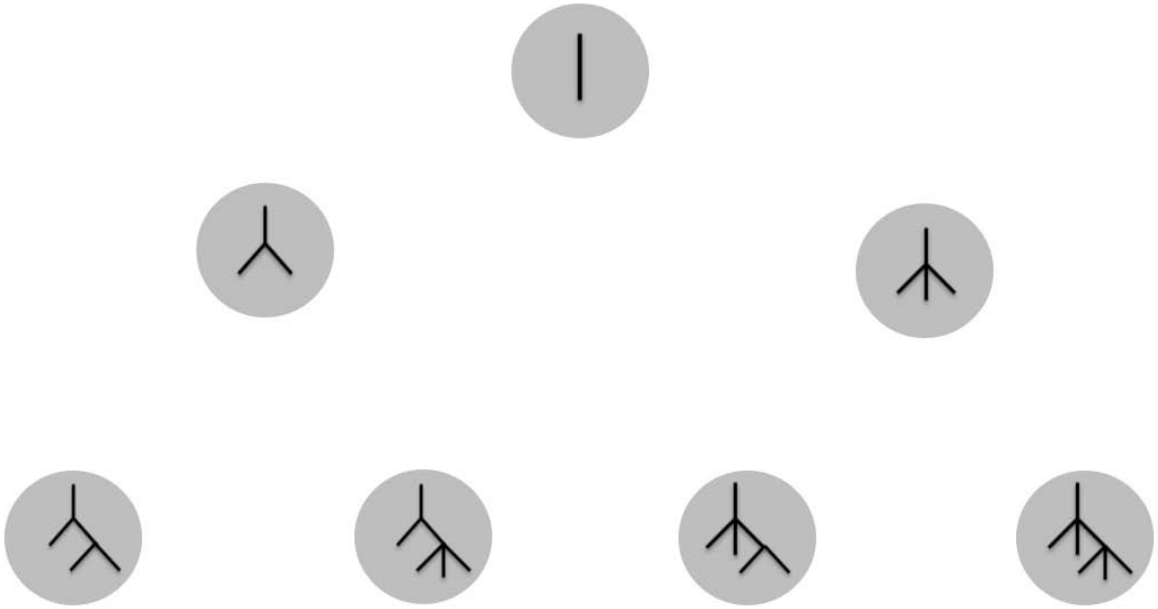
Limitations of Fraction Man Visualization

- Can become crowded and incomprehensible
- Hard to determine exact order of transitions as colors can be hardly distinguished
- Number of states shown on graph often does not match number of mathematical states
- Lack of structure: same state may appear at different locations in the graph when comparing multiple gameplays
- Goal of a level not easy to indentify

Concepts of Tree-based Visualization

- Incorporates ideas of network data visualization
- Based on “small multiples” where each state has a fixed location in the tree => Usable for comparison of gameplays of multiple students
- Identification of final goal via color
- Distinction between states reached and not reached via color
- Addition of a timeline to display order of transitions

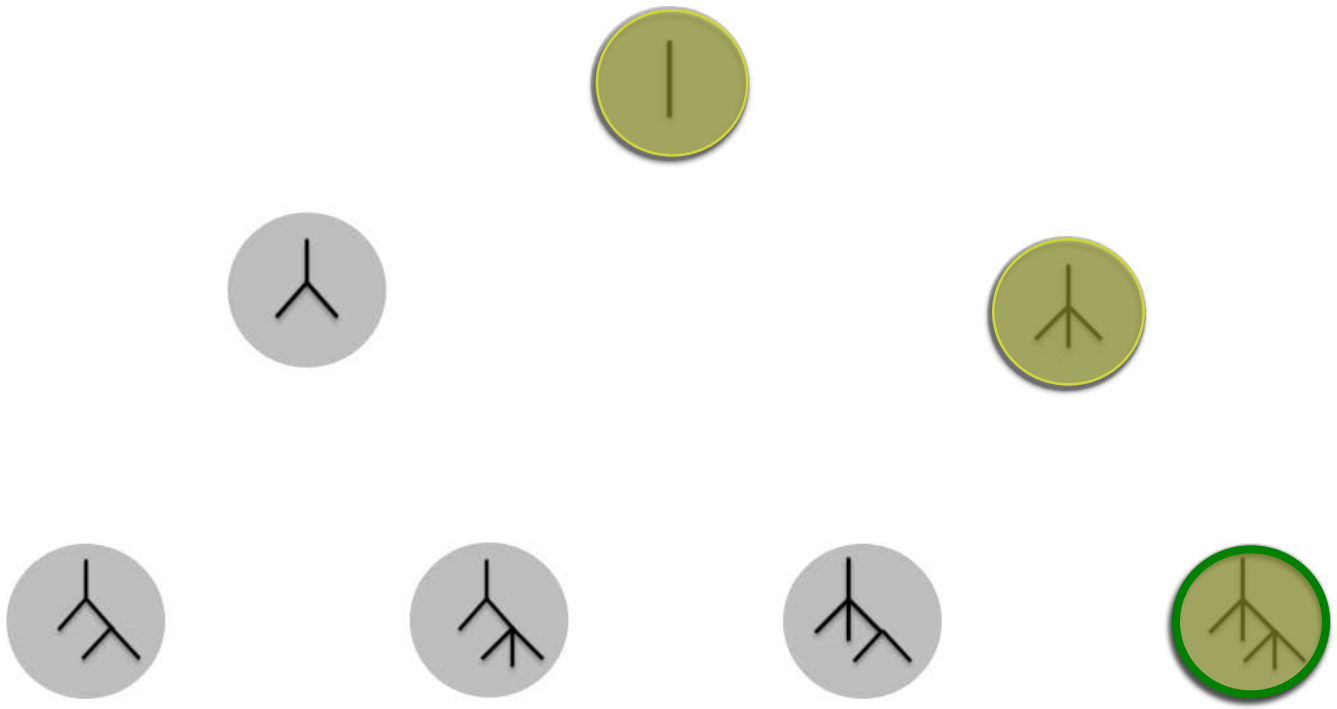
Initial “States”



“States”: Reached vs. Not Reached

States **reached** by the student

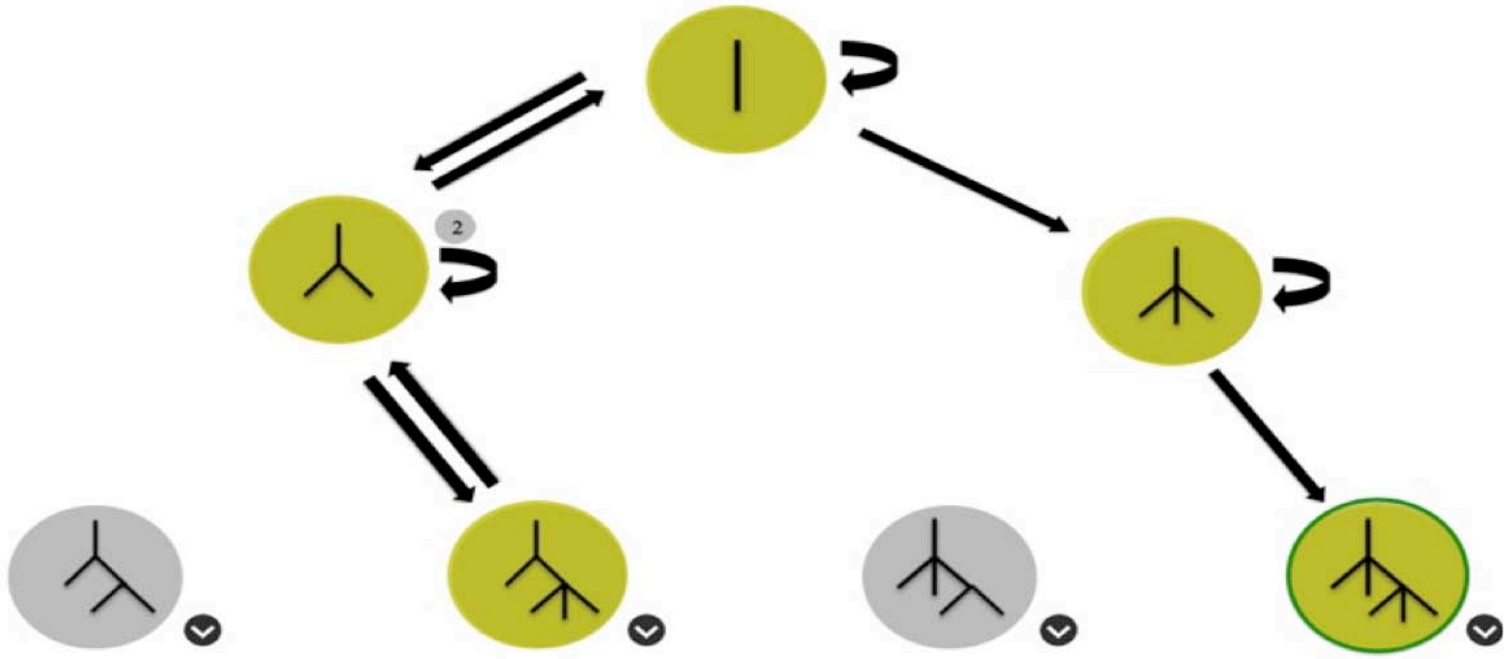
States **never reached** by the student



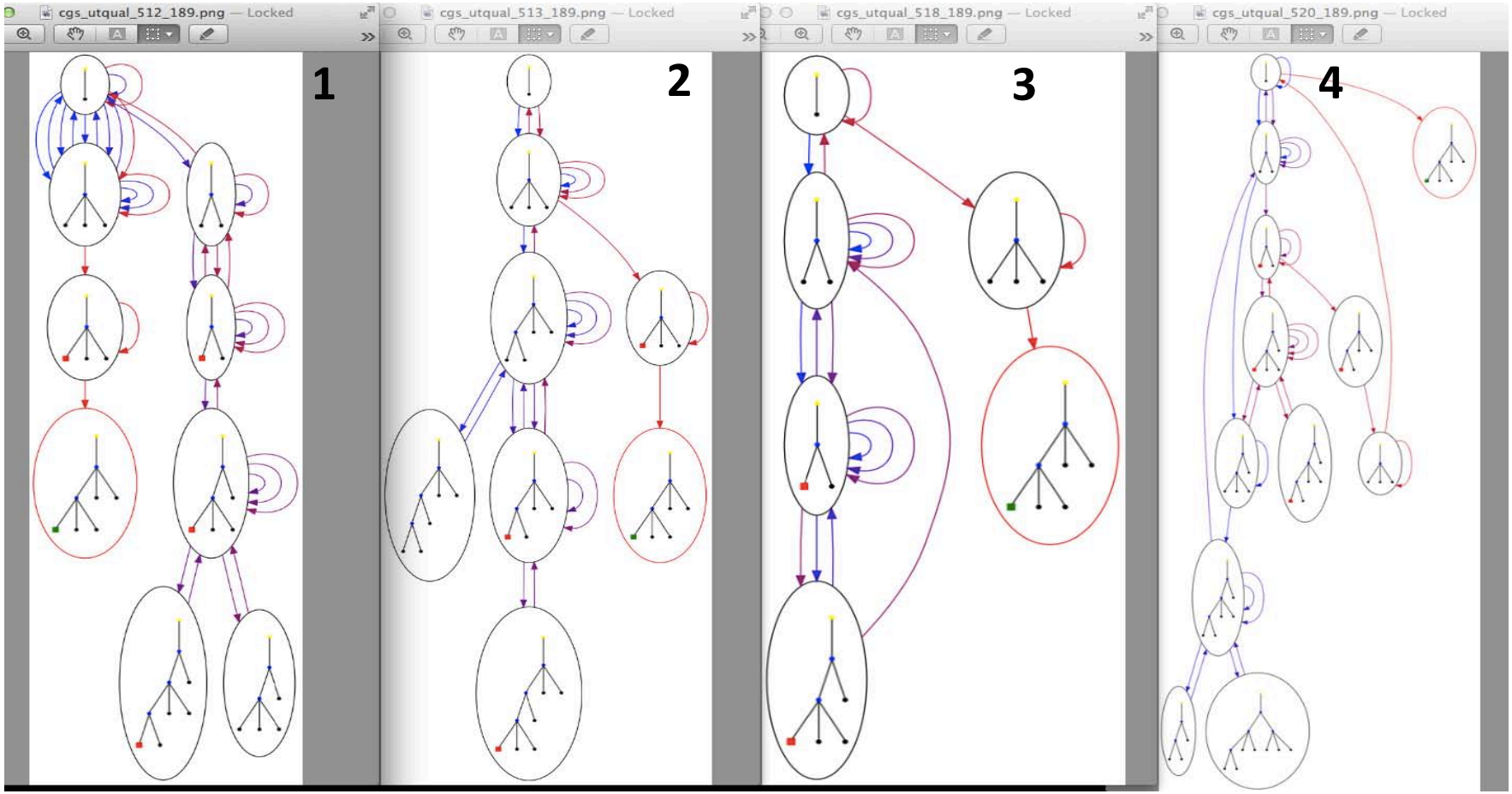
“States” & Transitions

States **reached** by the student

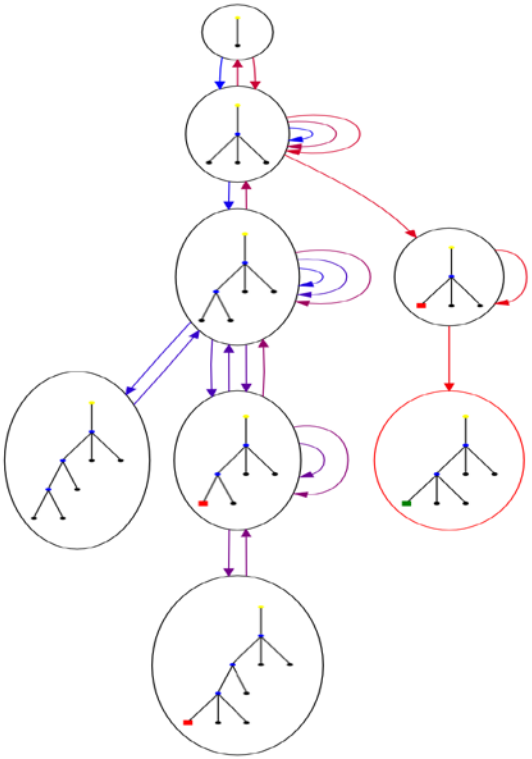
States **never reached** by the student

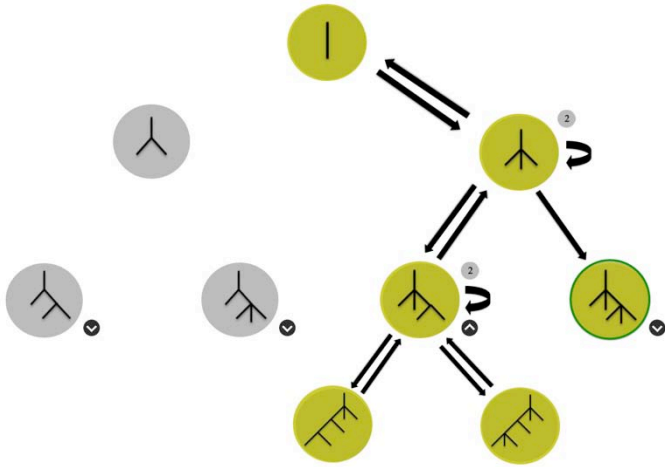


Level 189: Fraction Man Visualization

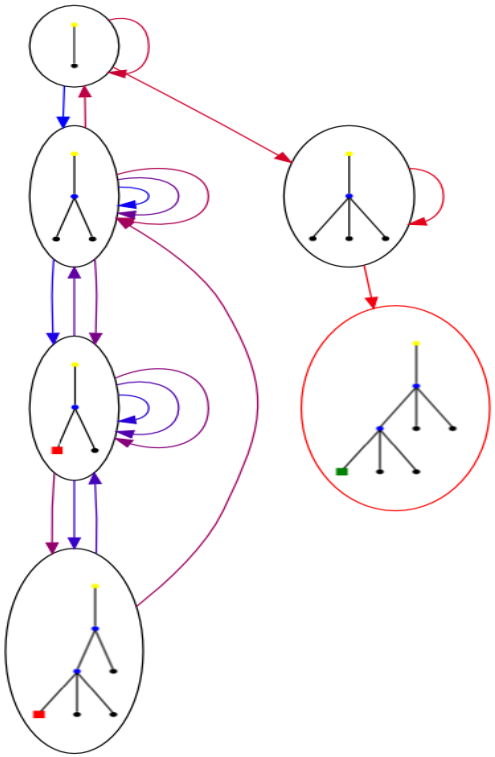


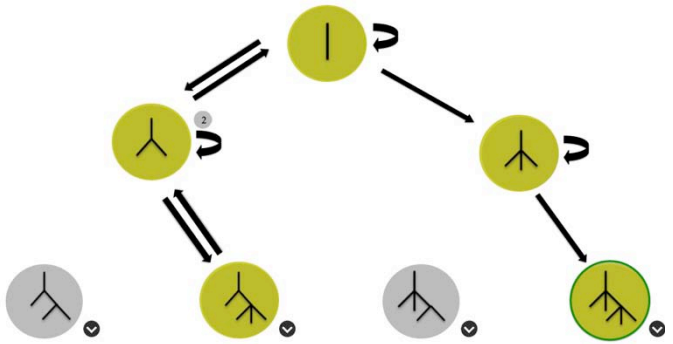
Student 513 Level 189



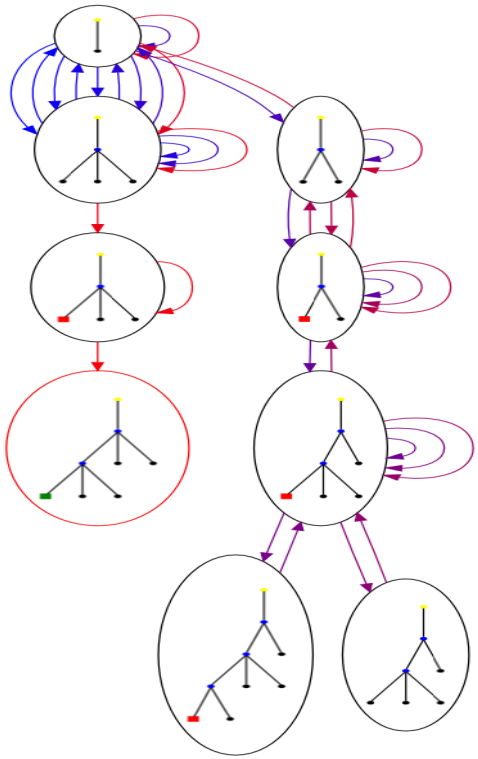


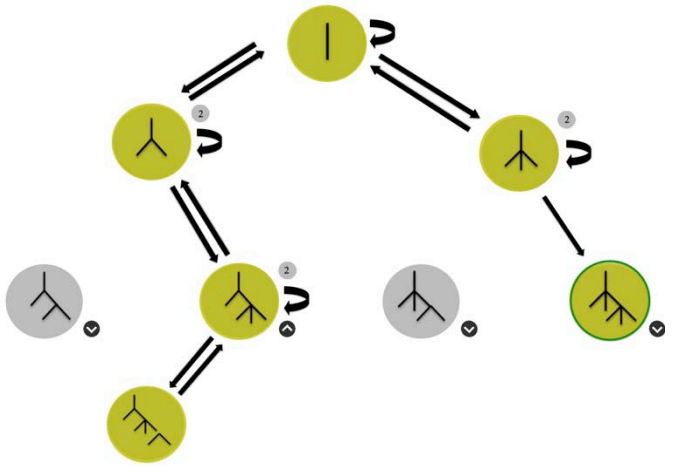
Student 518 Level 189



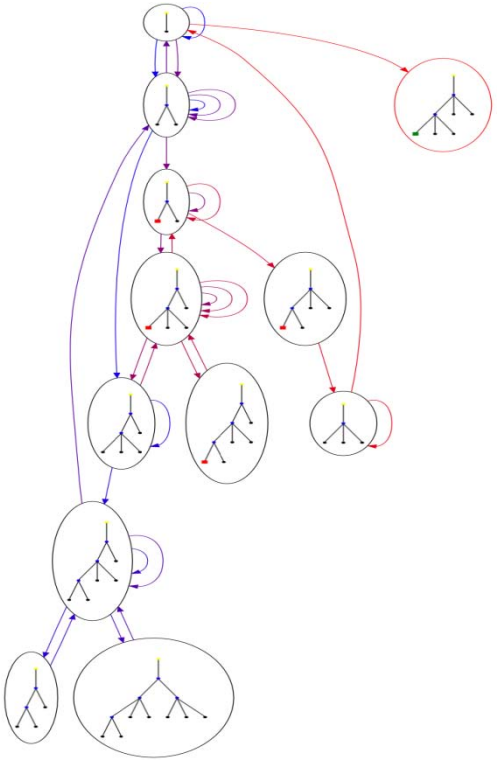


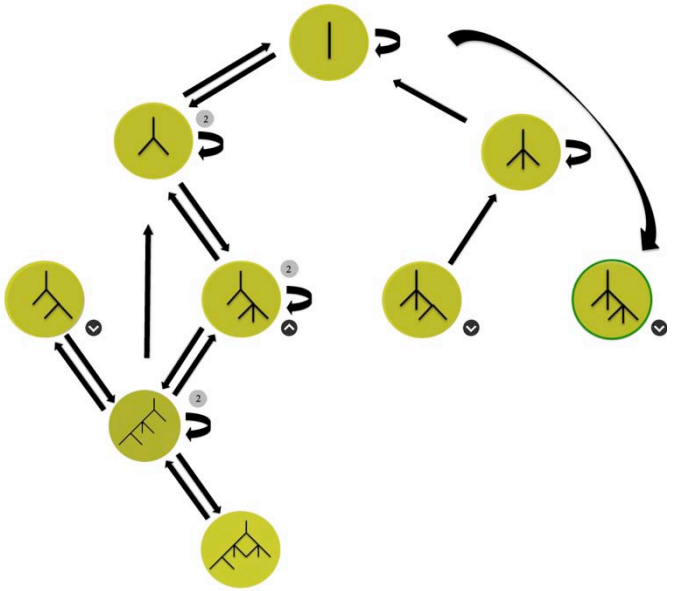
Student 512 Level 189



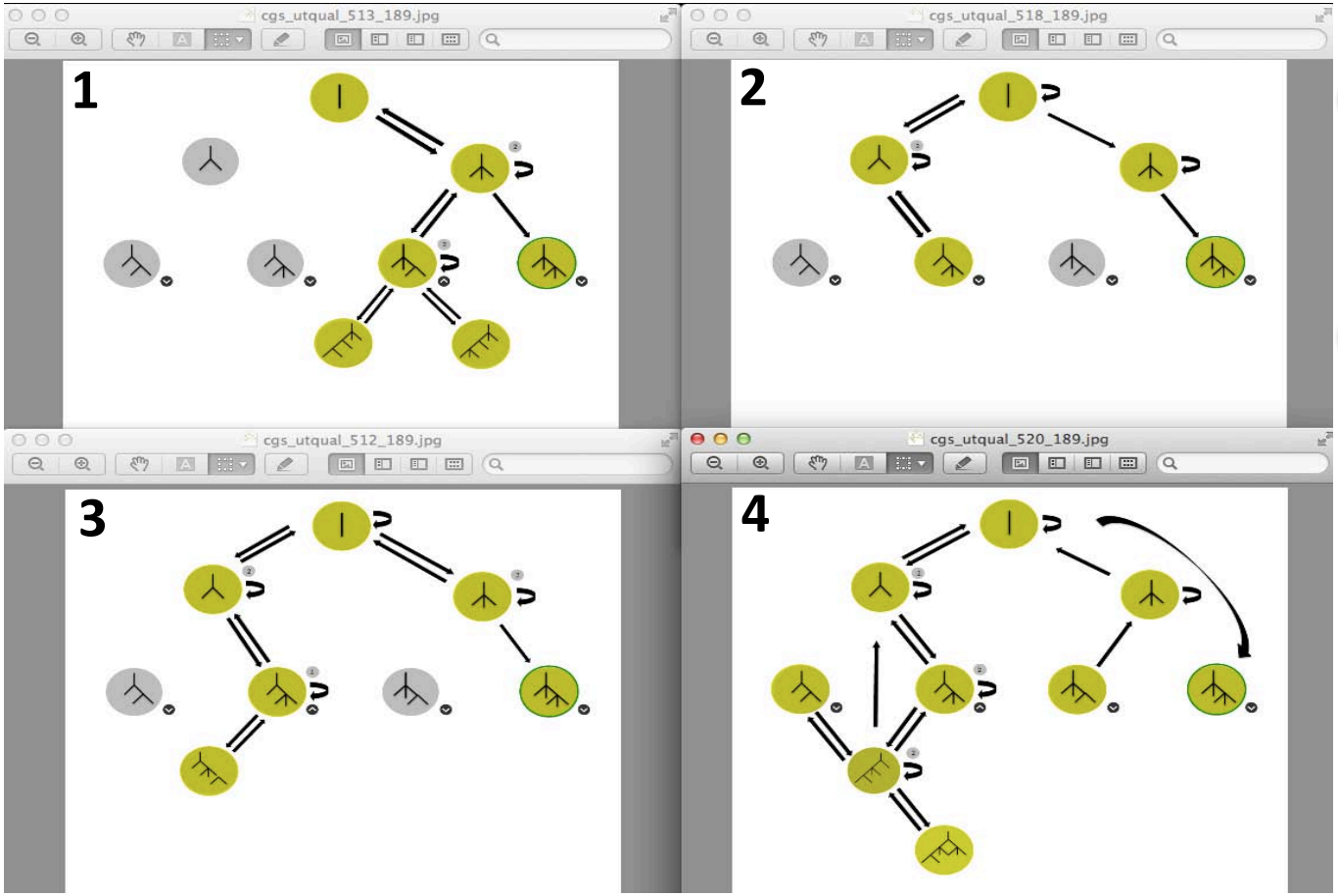


Student 520 Level 189



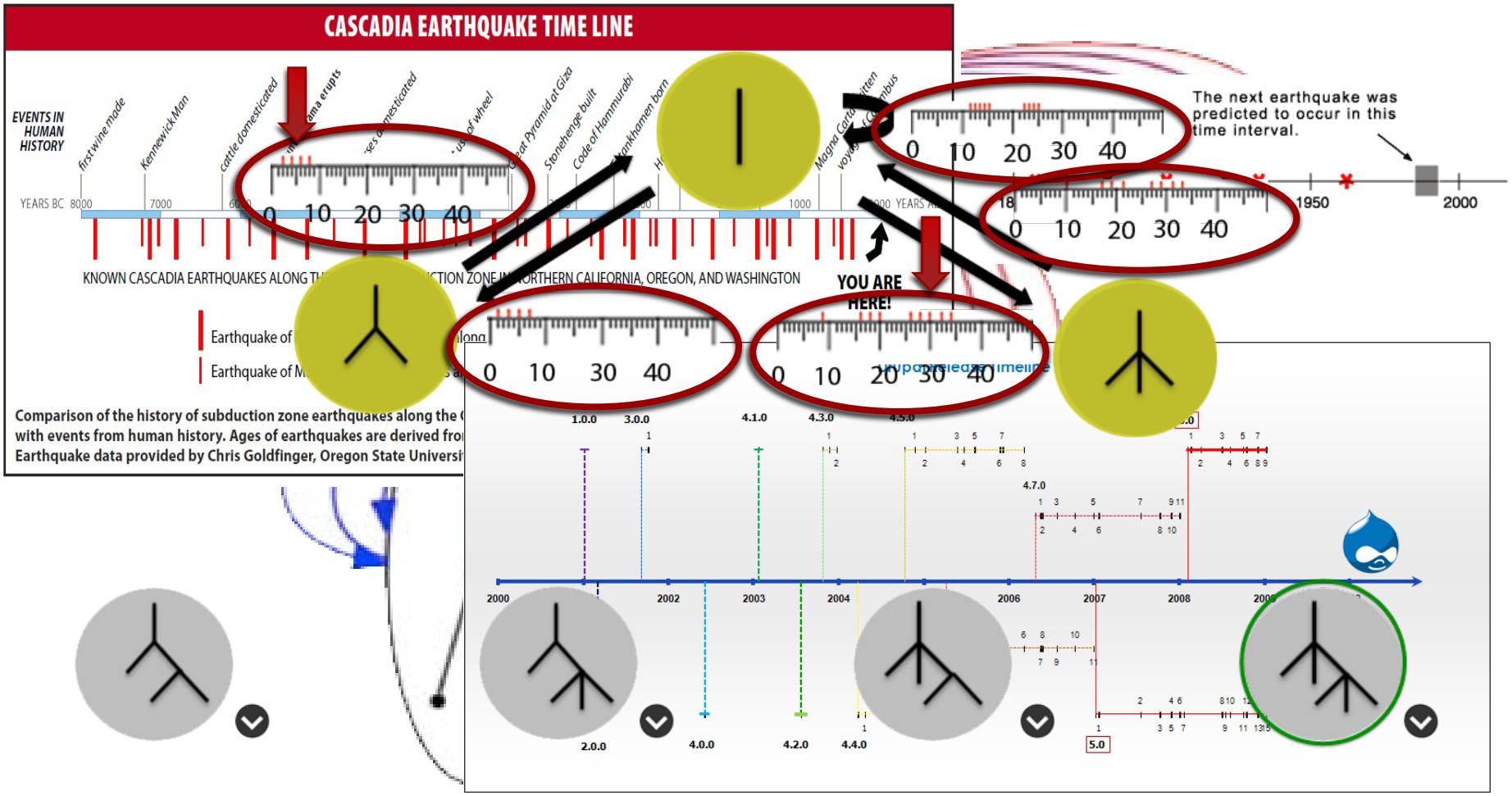


Level 189: New Visualization

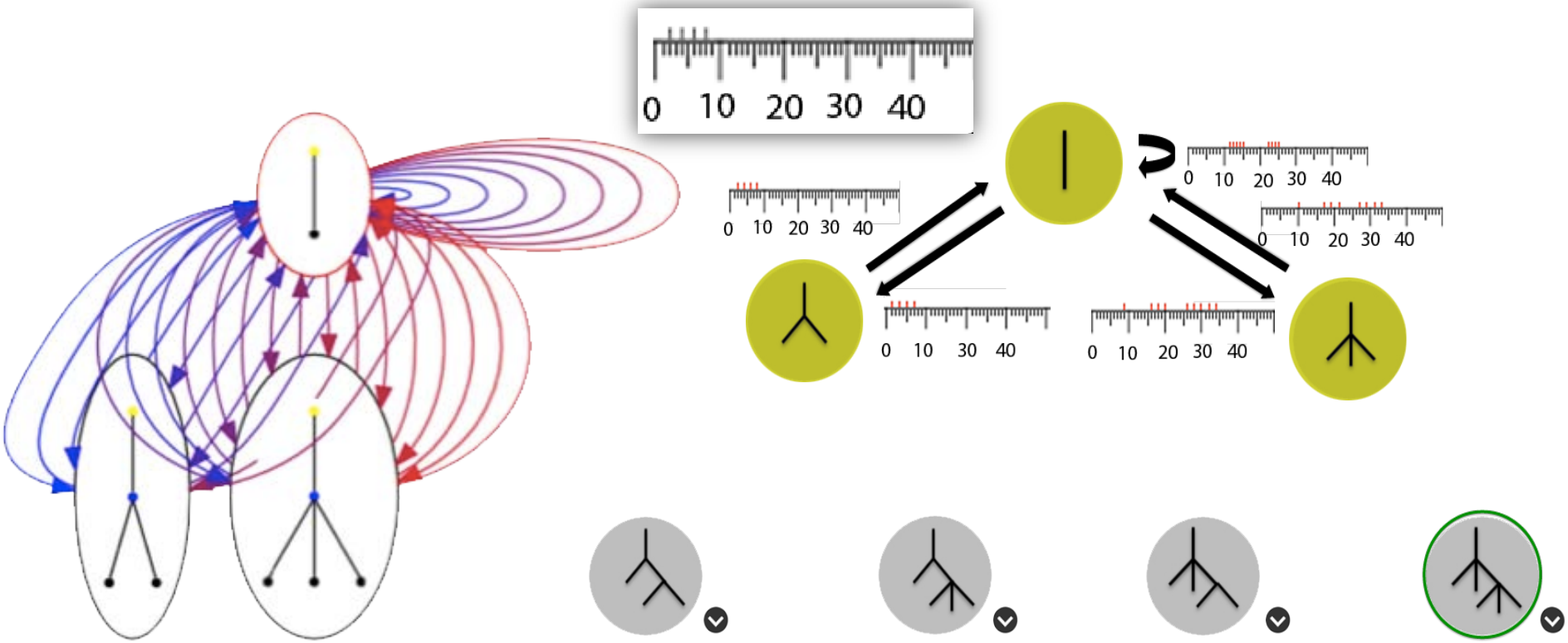


States **reached** by the student

States **never reached** by the student



The Alternative & the Timeline



Comparison

Left: Old visualization

Right: New visualization with timeline

Discussion

- Benefits:
  - Structured (small multiples!)
  - Easy to compare and contrast students' pathways through the same game level
  - Allows teachers/researchers to conduct exploratory analysis of student's pathways

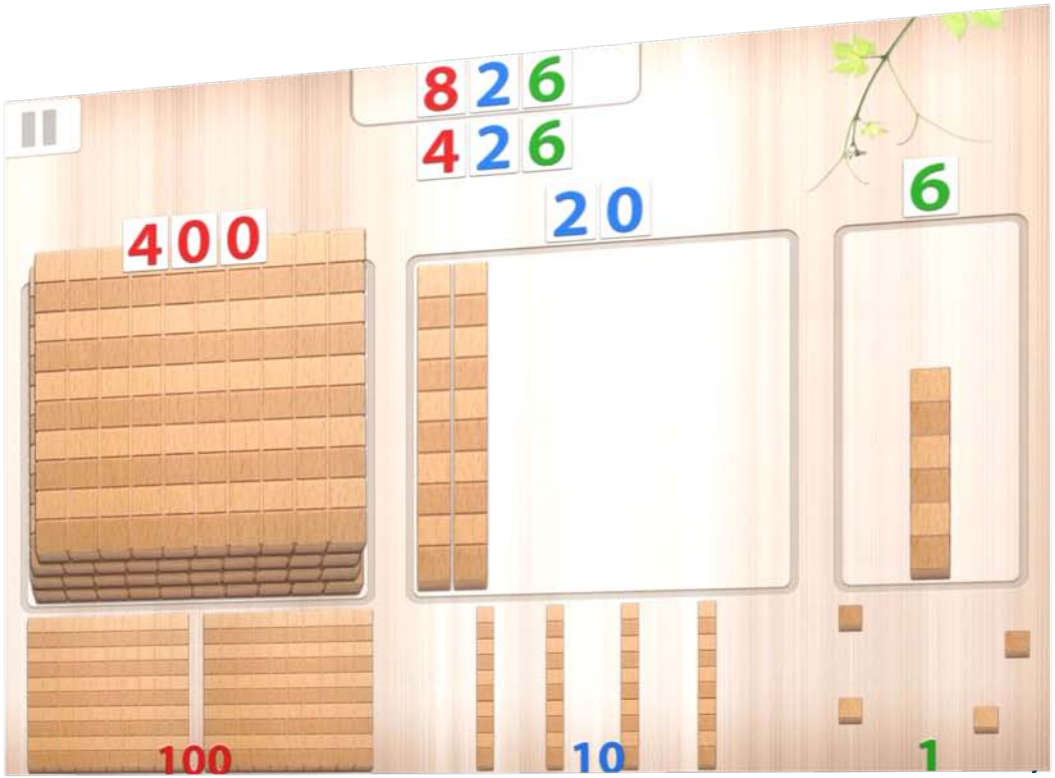
Case Study 2: An iPad Study

Published as:

Moyer-Packenham, P. S., et al. (2014): Developing Research Tools for Young Children's Interactions with Mathematics Apps on the iPad, 2014 Conference Proceedings, Hawaii International Conference on Education, Honolulu, Hawaii, USA.

Purpose of the Study

- The purpose of the project was to build theory and knowledge about the nature of young children's ways of thinking and interacting with virtual manipulative mathematics apps on the iPad.



Research Questions

Learning Performance and Efficiency:

- What are the immediate effects in learning performance and efficiency (pre vs. post) for children using virtual manipulatives for the iPad?

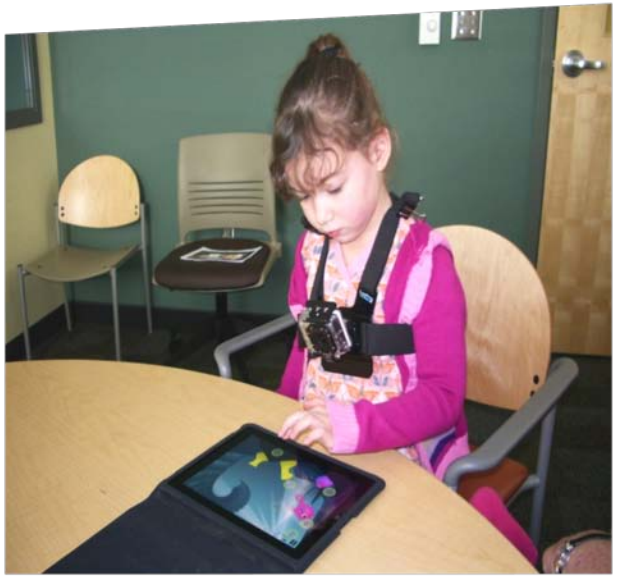
Learning Strategies:

- How do children interact with the virtual manipulatives on the iPad using the touch-screen capabilities?

Utah State University

Participants

- 100 children ages 3 to 8
  - 35 preschool, ages 3-4
  - 33 Kindergarten, ages 5-6
  - **32 Grade 2, ages 7-8**





- Demographic information were collected on age, gender, race, prior iPad use, etc.

Utah State University

Procedures: Clinical Interviews

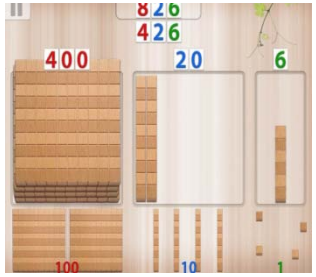
Sequence of the Interviews

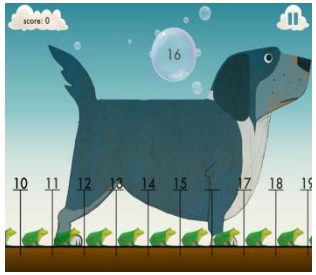
Interview	Grade Pre	Grade K	Grade 2
App #1 (pre)	Pink Tower: free moving	10-Frame	100s Chart
App #2 (learning)	Pink Tower: tapping	Hungry Guppy	Frog Number Line
App #3 (learning)	Red Rods	Fingu	Counting Beads
App #1 (post)	Pink Tower: free moving	10-Frame	100s Chart
App #4 (pre)	Base-10 Blocks	Base-10 Blocks	Base-10 Blocks
App #5 (learning)	Base-10 Blocks: 1-5	Base-10 Blocks: 11-20	Zoom Number Line
App #6 (learning)	Base-10 Blocks: numerals	Base-10 Blocks: numerals	Place Value Cards
App #4 (post)	Base-10 Blocks	Base-10 Blocks	Base-10 Blocks

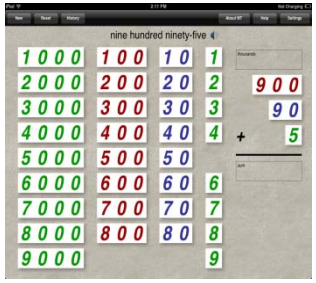
Grade 2 Interview Apps

Quantities

- Pre/Post
  - Base-10 Blocks
- Activity A
  - Math Motion Zoom (Levels 2-4)
- Activity B
  - Place Value Cards (3-digit problems without zeros)

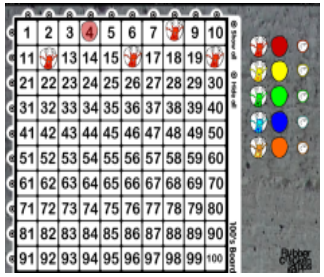


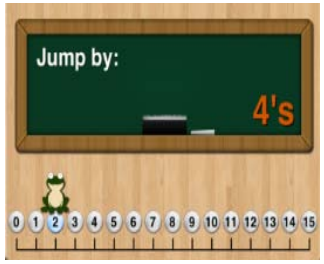


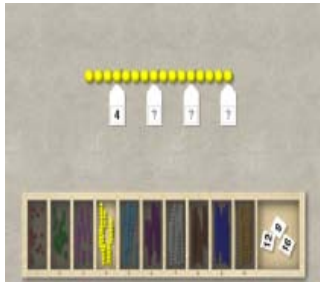


Skip Counting

- Pre/Post
  - 100s Chart
- Activity A
  - Number Lines (Skip Counting Tool
- Activity B
  - Skip Counting Beads













Data Collection

Video Cameras:

- Wall-mounted
- Go-Pro

Utah State University

Data Analysis

- Base-10 Blocks: 6 performance variables (1: Model a number between 12 and 30; 2: Model a number between 54 and 62; 3: Model 181; 4: Model a number between 181 and 200; 5: Model 267; & 6: Model a number 20 less than 267)
- 100s Chart: 3 performance variables (skip counting by 4, 6 & 9)
- Performance data were scaled from 0 (very poor) to 1 (excellent)
- Analyzing learning performance from Pre- and Post-assessments

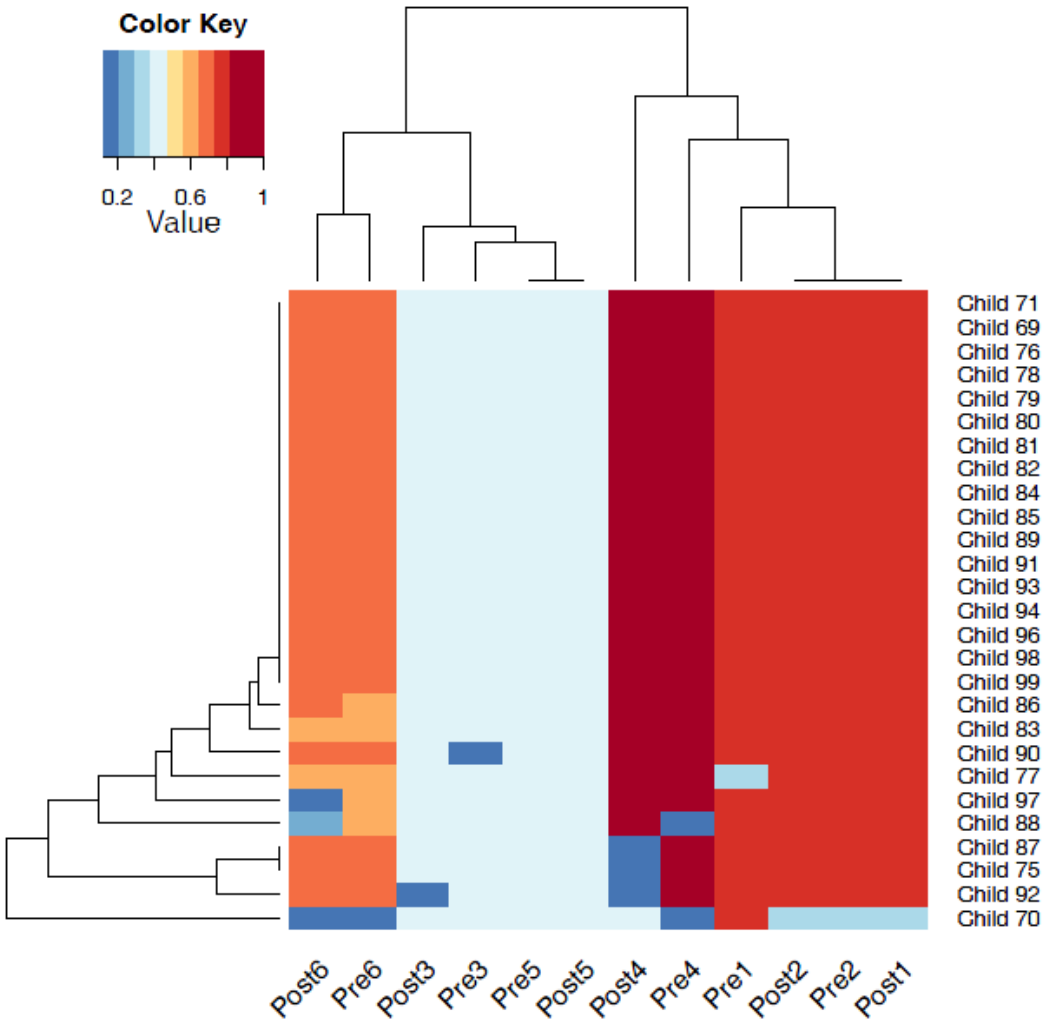
UtahState University

Grade 2: Base-10 Blocks



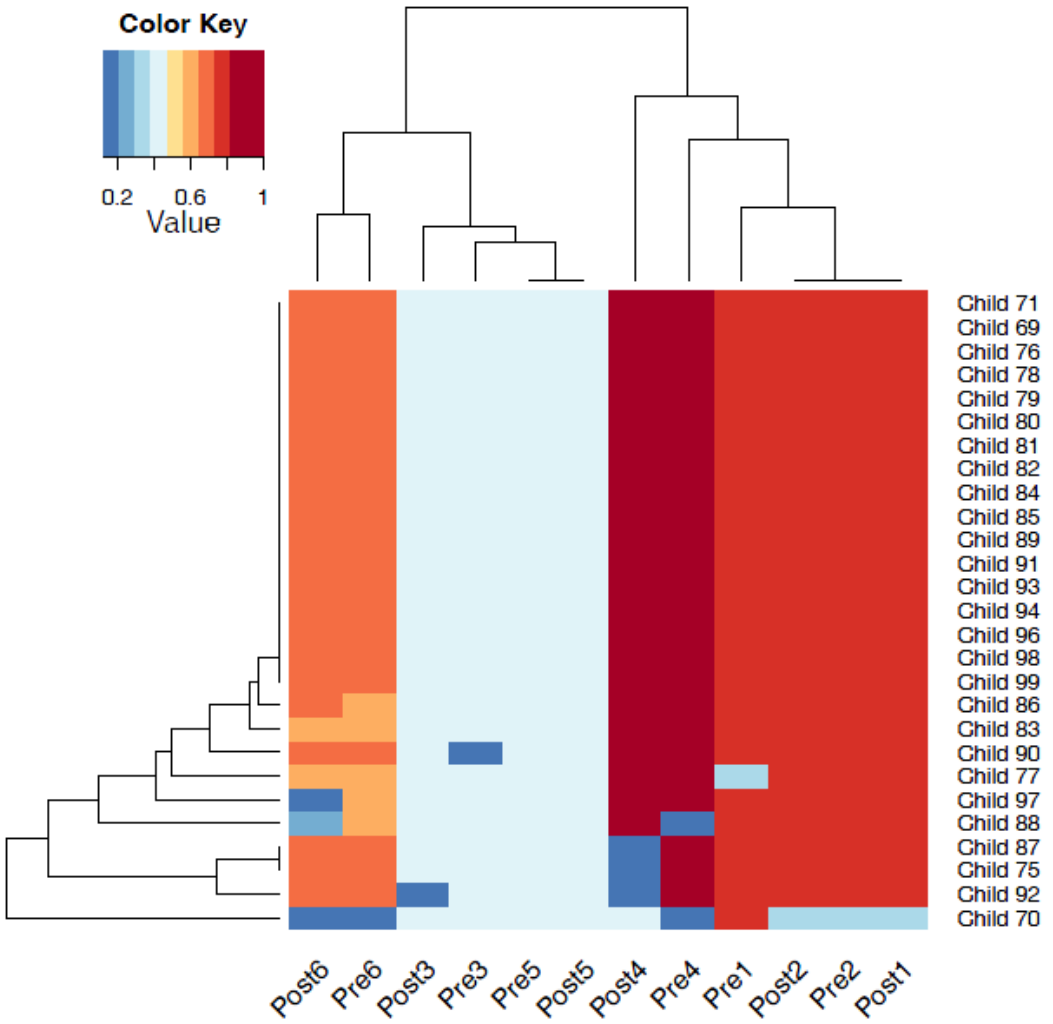
Base-10 Blocks (1)

- 17 of the 27 children had totally identical outcomes. Moreover, the outcomes for the two children # 75 and 87 were totally identical. Only eight children were more different. Child # 70 was most unusual with only a single score higher than 0.5.
- Scores from the six Pre and Post test pairs matched each other very closely.
- Variables Post1, Pre2, and Post2 had totally identical outcomes with a score of 0.875 for all but one child. Variable Pre1 matched them for all but two children.



Base-10 Blocks (2)

- Variables Pre5 and Post5 had totally identical outcomes with a score of 0.5 for all children.
- Variables Pre3 and Post3 matched them for all but two children.
- Variables Pre6 and Post6 had almost identical outcomes with a score of 0.75 for most children. Only three children's outcomes were different in their pre/post tests.
- Variables Pre4 and Post4 had almost identical outcomes with a score of 1.0 for most children. Only five children's outcomes were different in their pre/post tests.



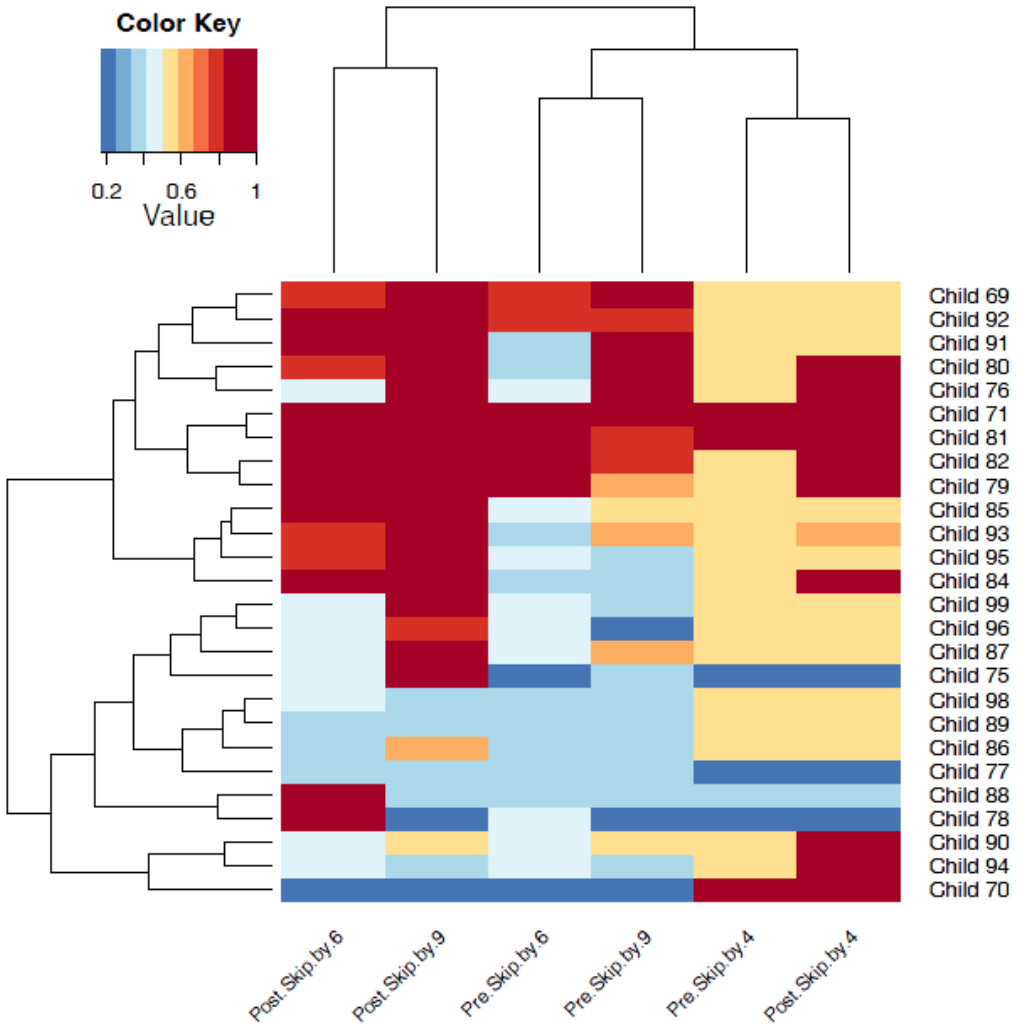
Grade 2: Skip Counting

100s Chart



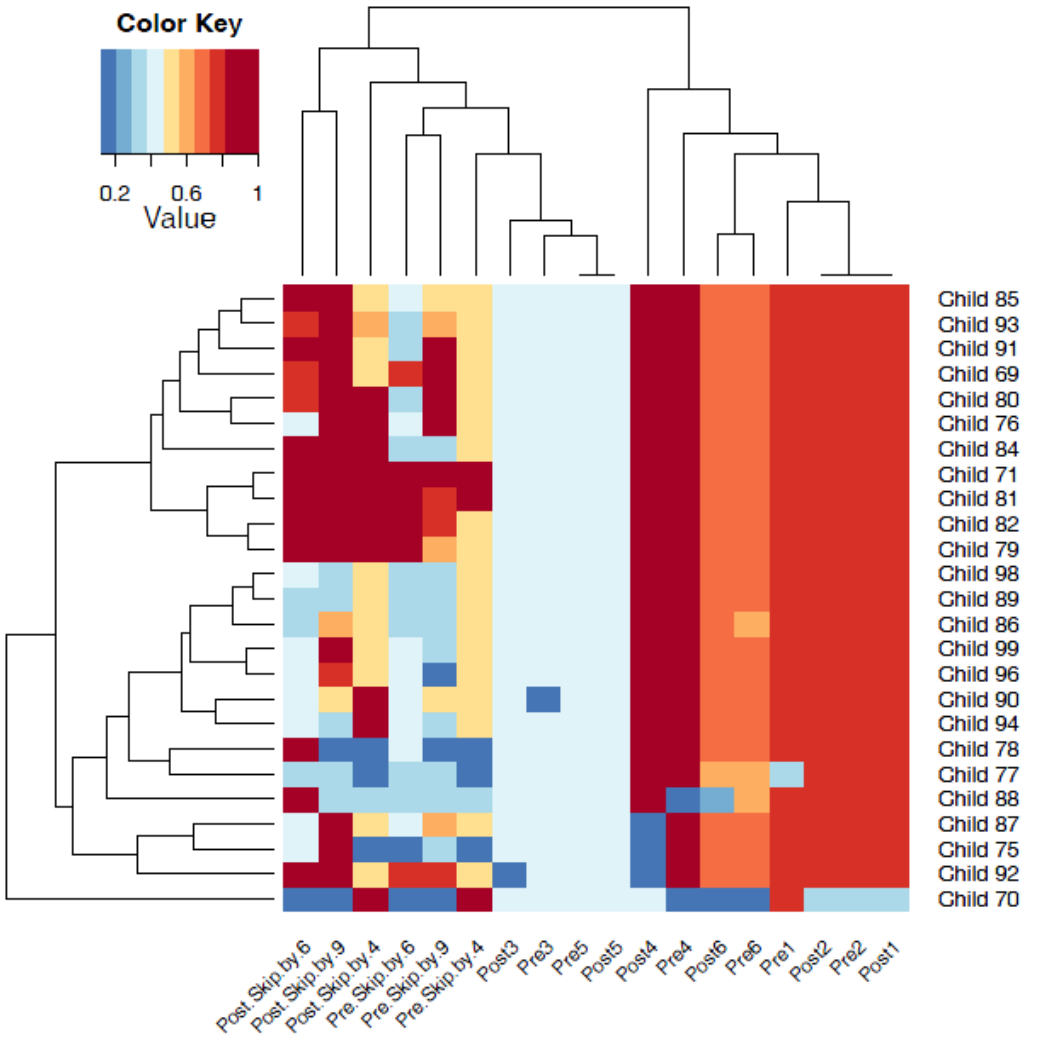
100s Chart

- In the dendrogram, there are three pairs of two variables: Pre.Skip.by.4 and Post.Skip.by.4, Pre.Skip.by.6 and Pre.Skip.by.9, and Post.Skip.by.6 and Post.Skip.by.9.
- There are no children who had exactly the same outcomes.
- There are two main clusters: In cluster 1 (13 children: # 69 to # 84 on the right), children had medium to high scores for most of the variables. In cluster 2 (13 children: # 99 to # 70 on the right), children had low to medium scores for most of the variables.
- Child # 71 is the only child with a perfect 1.0 for all six variables.



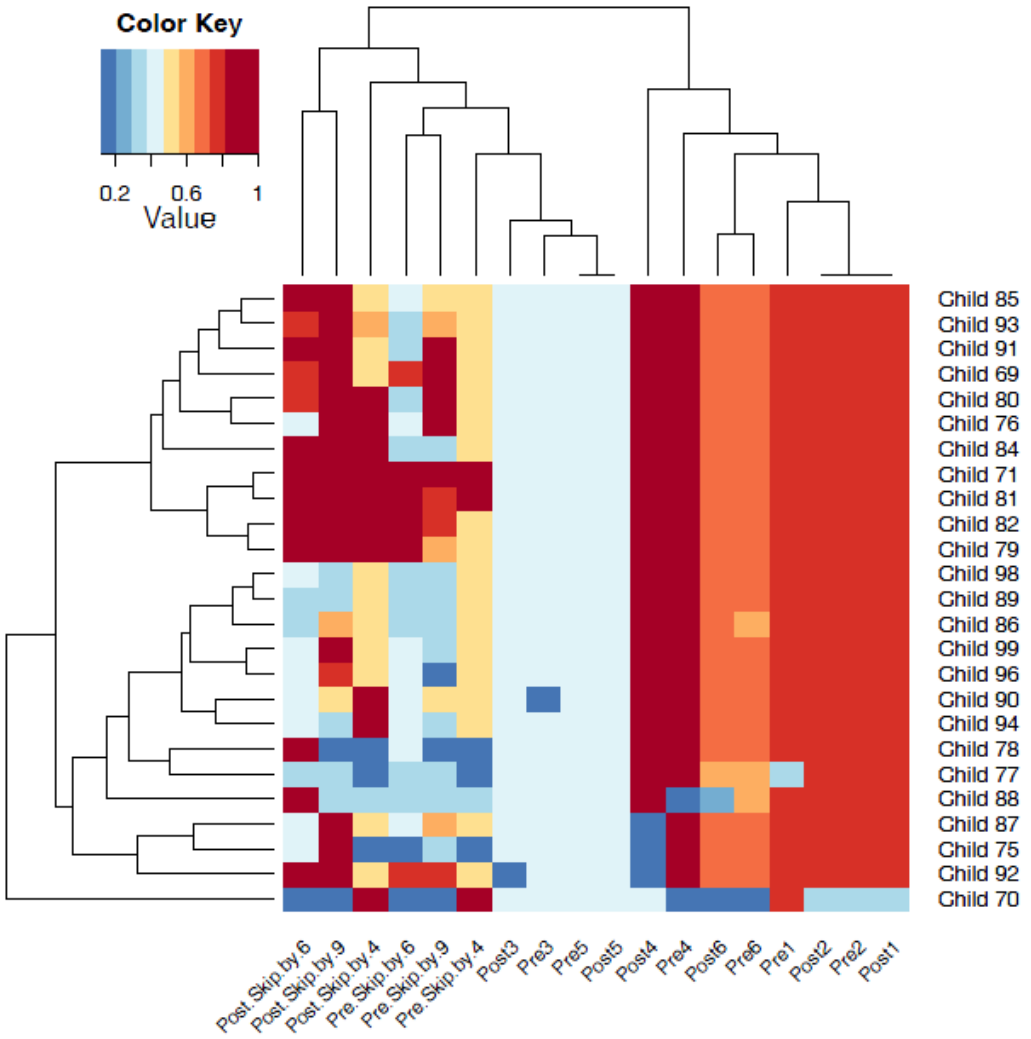
Summary (1)

- In the dendrogram, there are two main groups of variables: Pre/Post 1, 2, 4, and 6 (group 1) and Pre/Post 3 and 5, and Pre/Post skip.by 4, 6, and 9 (group 2). This means Pre/Post 3 and 5 are somewhat closer to the variables from 100 Chart than to the eight other variables from Base 10.
- There are no children who had exactly the same outcomes.



Summary (2)

- There are two main clusters: In cluster 1 (11 children: # 85 to # 79 on the right), children had medium to high scores for most of the variables. In cluster 2 (13 children: # 98 to # 92 on the right), children had low to medium scores for most of the variables.
- Child # 70 is the most unusual child that differs considerably from both main clusters of children.
- Second-grade children increased their performance on the skip counting app, but not on the base 10 app.



Conclusions

- Visual approach effective to see unexpected structure in data
- Combination of different techniques most effective
- Can be used for almost all types of data, including environmental and medical data
- Stay tuned for VDM techniques for spatial data ☺

Questions ???