Peeking through the Classroom Window : A Detailed Data-Driven Analysis on the Usage of a Curriculum Integrated Math Game in Authentic Classrooms

Preya Shabrina[a]
Presenter

Ruth Okoilu Akintunde[a]

Mehak Maniktala[a]

Dr. Tiffany Barnes[a]

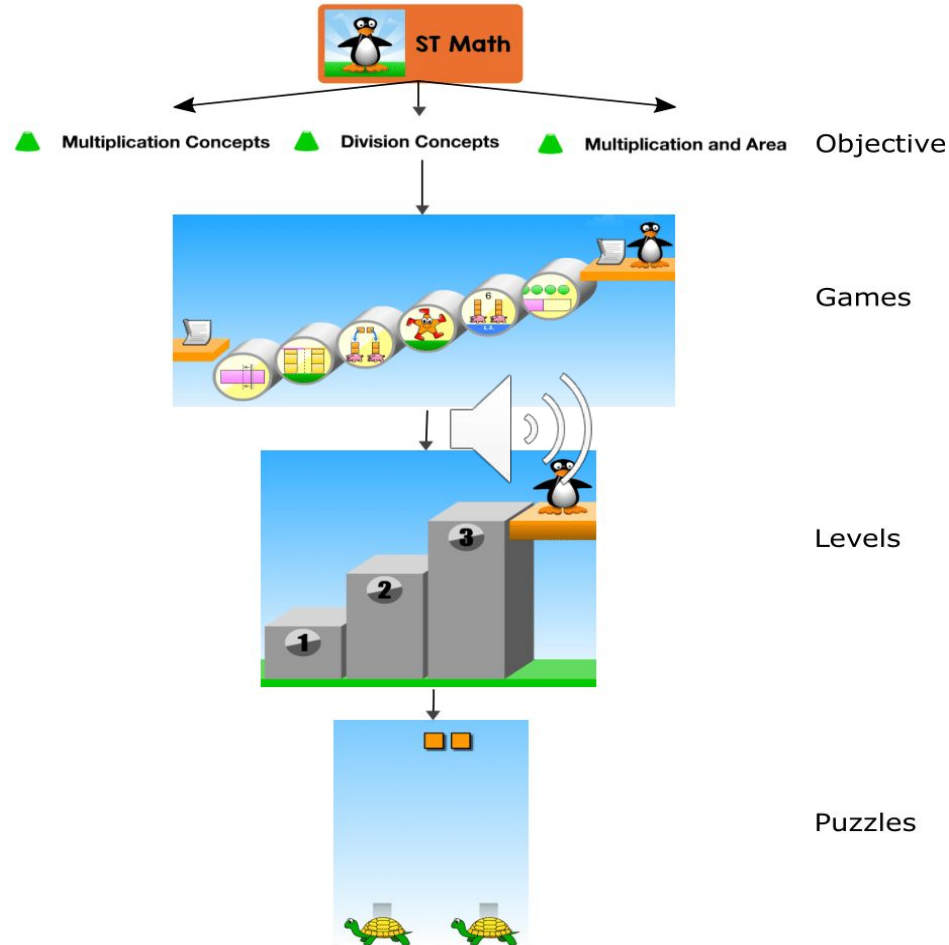**Dr. Collin Lynch[a]**

Dr. Teomara Rutherford[b]

a - Department of Computer Science, North Carolina State University
b - Education & Human Development, University of Delaware

Spatial Temporal Math (STMath) - Developed by MIND Research Institute

- Currently used by 1.2 million students and 56,000 teachers across 48 states in the USA

- Curriculum Integrated - contents are designed and organized to align common core and state level mathematics curriculum

- Uses spatial puzzles and restricts the use of textual instruction

- Organized as a hierarchy of Objectives, Games, Levels and Puzzles

Spatial Temporal Math Hierarchy

ST Math

Multiplication Concepts   Division Concepts   Multiplication and Area   Objective

Games

Levels

Puzzles

Motivation

- Peddycord Liu's 2019 field study on Teaching Practices surrounding ST Math

    - **Eight Female Teachers from 6 schools (Teacher Selection Bias)**
    - **Time of the Study (at the end of the academic year)**

To get a generalized idea on STMath usage, they pointed out the necessity of -

- **Field Studies at different times of a school year**
- **With non-volunteer teachers!!!**

Prior Research

- Field study to identify the association of classroom structures with students' learning attitude and performance. [Ke, 2006] [1]
  - **found that cooperative goal structure classrooms where peers cooperate with each other are most beneficial**

- Hangh et al. (2010) investigated how teachers used educational games in classrooms [3].
  - **teachers shift between roles to meet the demands of different game modalities and situations during a gameplay session**

- Another Investigation on how teachers integrate games into instruction and the associations between teaching practices and student achievement. [Callaghan et al, 2018] [2]
  - **school differences may contribute to both teaching practices and student outcomes**

Research Questions

1. Can we identify when students were playing together in classrooms (game play sessions) from their sequential gameplay data?

2. If yes, can we figure out what was the classroom setup, how teachers used the game, and the year-round usage?

3. What are the features that varied across game play sessions?

4. What is their relationship with session performance?

Data

STMath Students' Sequential Game Play Data

- **One transaction for each level attempt**
- Each row contains -
  - Student ID
  - Timestamp indicating level attempt completion
  - Login place (home or lab)
  - Objective, game, and level ID
  - number of puzzles passed, and number of total puzzles in that level

Playing a Level in STMath

Data

## STMath Students' Sequential Game Play Data

- One transaction for each level attempt
- Each row contains -
    - Student ID
    - Timestamp indicating level attempt completion
    - Login place (home or lab)
    - Objective, game, and level ID
    - number of puzzles passed, and number of total puzzles in that level

  - Performance = puzzles passed / number of total puzzles in that level

Gameplay Session Identification

1. Grouping level-attempts that are in the same gameplay session

2. Assigning them with a session number

Gameplay Session Identification
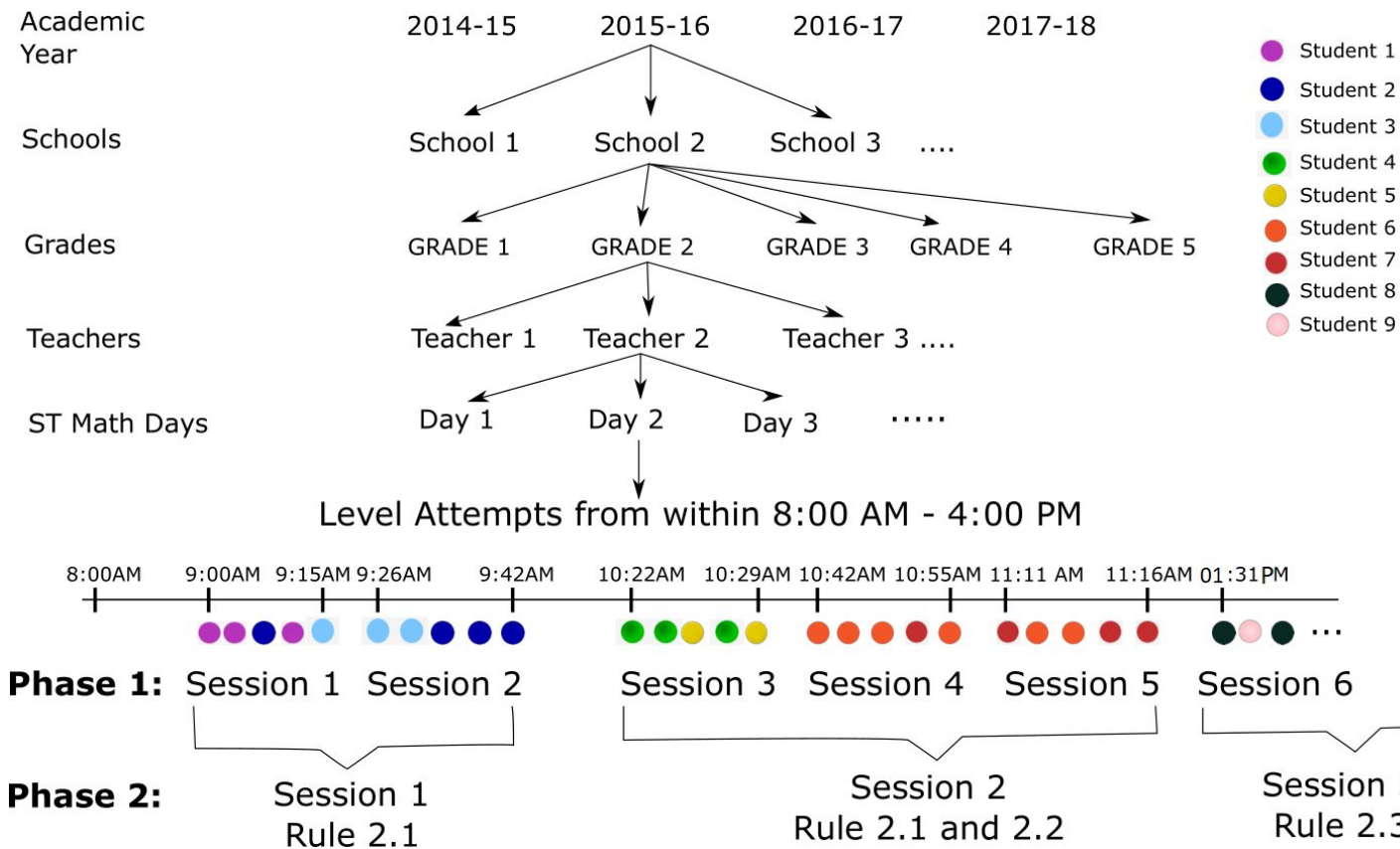
Identified Gameplay Sessions

- Identified **586,026 sessions**
    - Analyzed Sessions that are **45 minutes long and 5-25 students**
    - **98,309 sessions (21% of total sessions)**

Components of Classroom Integration/Session Features

Feature	Mean	SD
Session length (min)[2]	**28.05**	**9.64**
Class size[2]	**9.43**	**4.25**
Avg. # of level attempts[1]	**5.35**	**2.49**
Avg. min. played/student[1]	**7.78**	**5.96**
Max. % of class[2]	**0.77**	**0.2**
Start time variance (min.)[2]	77.48	**71.23**
Finish time variance (min.)[2]	63.56	**69.27**
# previous ST Math days	.	.
# previous sessions	.	.
% students practised	**0.13**	**0.26**

[1] Cutoff at 3 standard deviations   [2] Cutoff using Interquartile range

Components of Classroom Integration/Session Features

Binary Features	No	Yes
Disjoint small group?	78,341	**5,317**
Before-session practise?	43,779	39,879

start_time_variance + finish_time_variance + Disjoint small group
=> Class Format/organization of students during the gameplay session

Selected Sessions

After Outlier Removal -

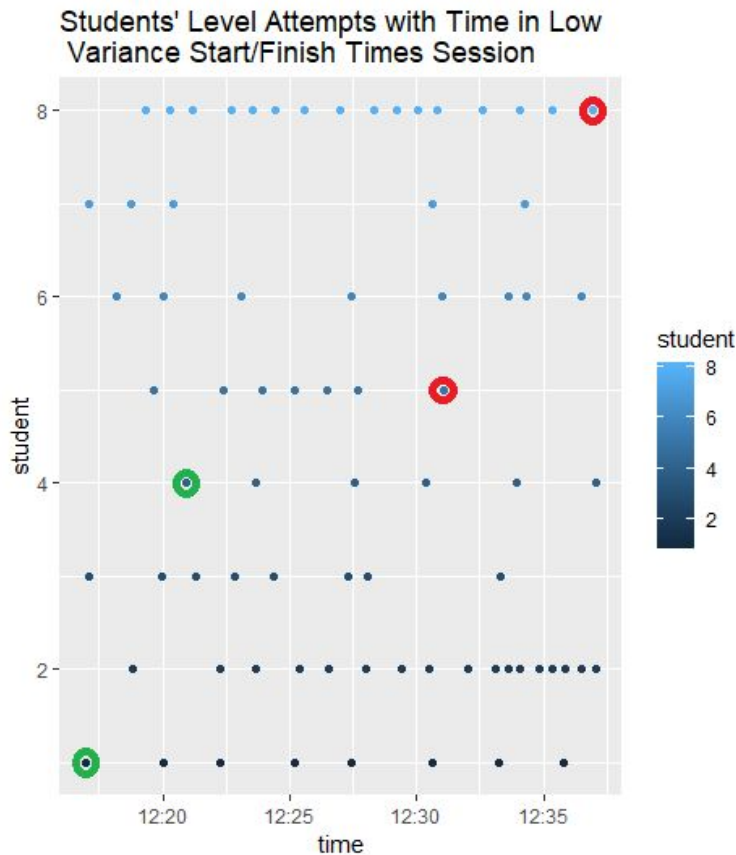- Finally, we had **83,658 Sessions**

Study 1 : Class Formats in Gameplay Sessions, Gameplay Sessions from Teachers' Perspectives and Year-Round usage of ST Math

Peddycord Liu et al. observed -
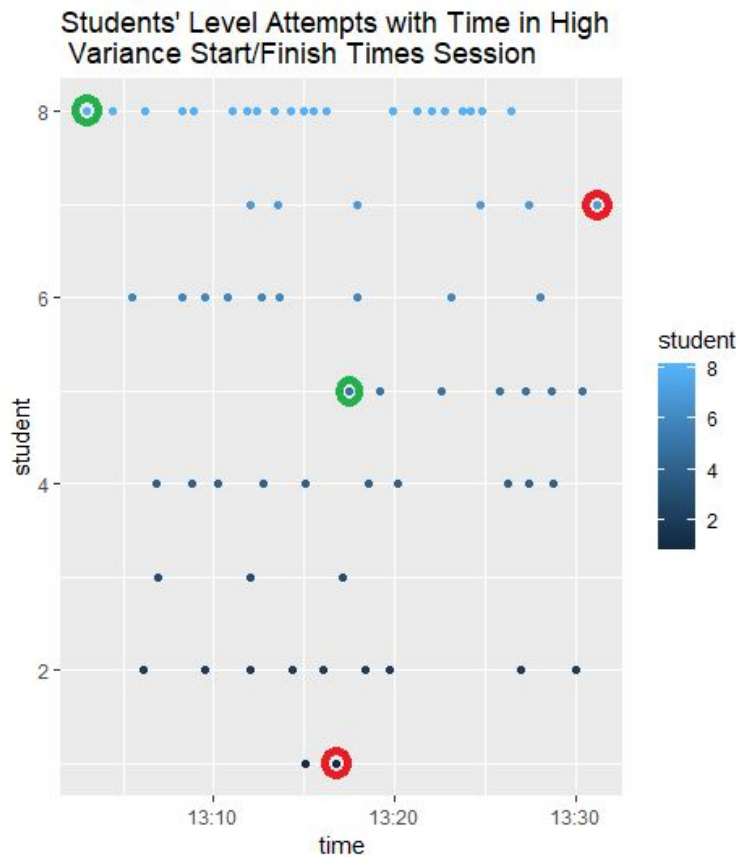
Three Class Formats

Observed Class Formats - Low Variance Start/Finish Times Format

Students' Level Attempts with Time in Low Variance Start/Finish Times Session

- Based on **Lab-Seating**
- **Start and End Times Vary No More than 5 Minutes**

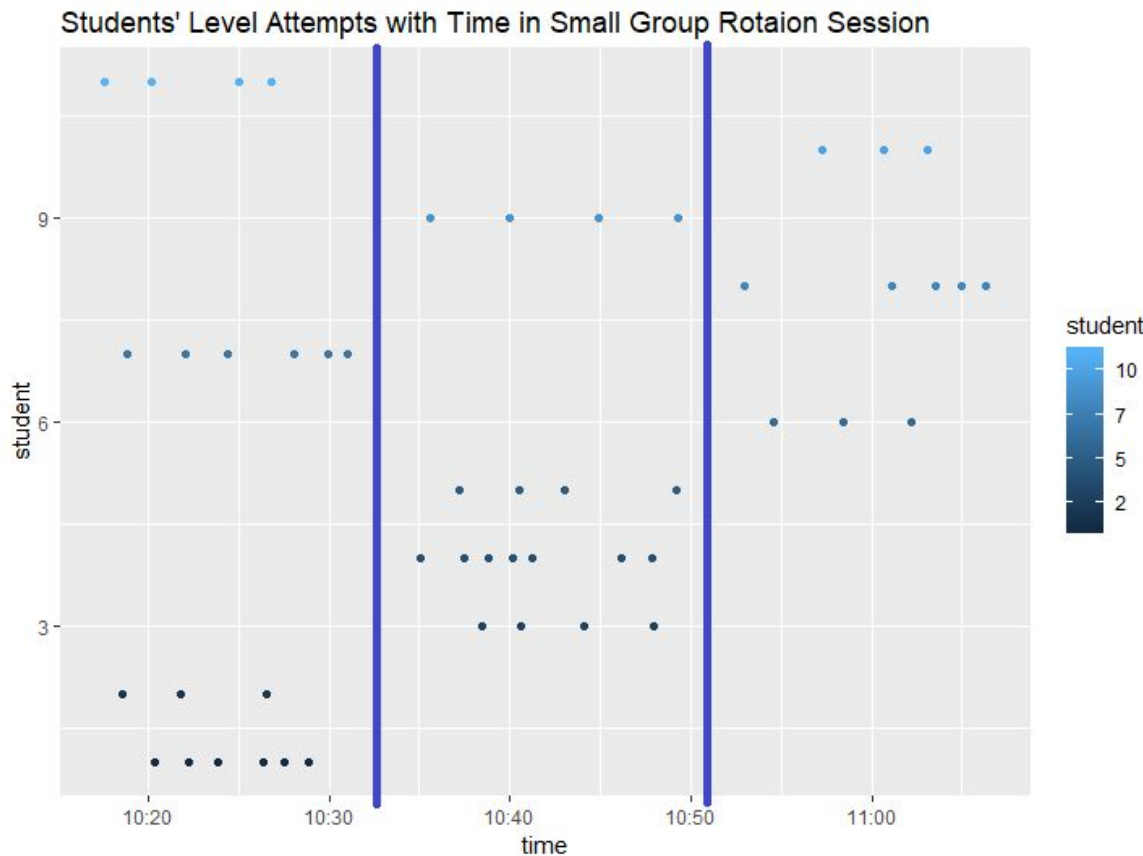# Observed Class Formats - High Variance Start/Finish Times Format

Students' Level Attempts with Time in High Variance Start/Finish Times Session

- Based on **Free-Seating**
- **Start and End Times Vary More than 5 Minutes**

Observed Class Formats - Small Group Rotation Format

Students' Level Attempts with Time in Small Group Rotaion Session

- Based on **Rotation-Seating**

Class Formats Summary

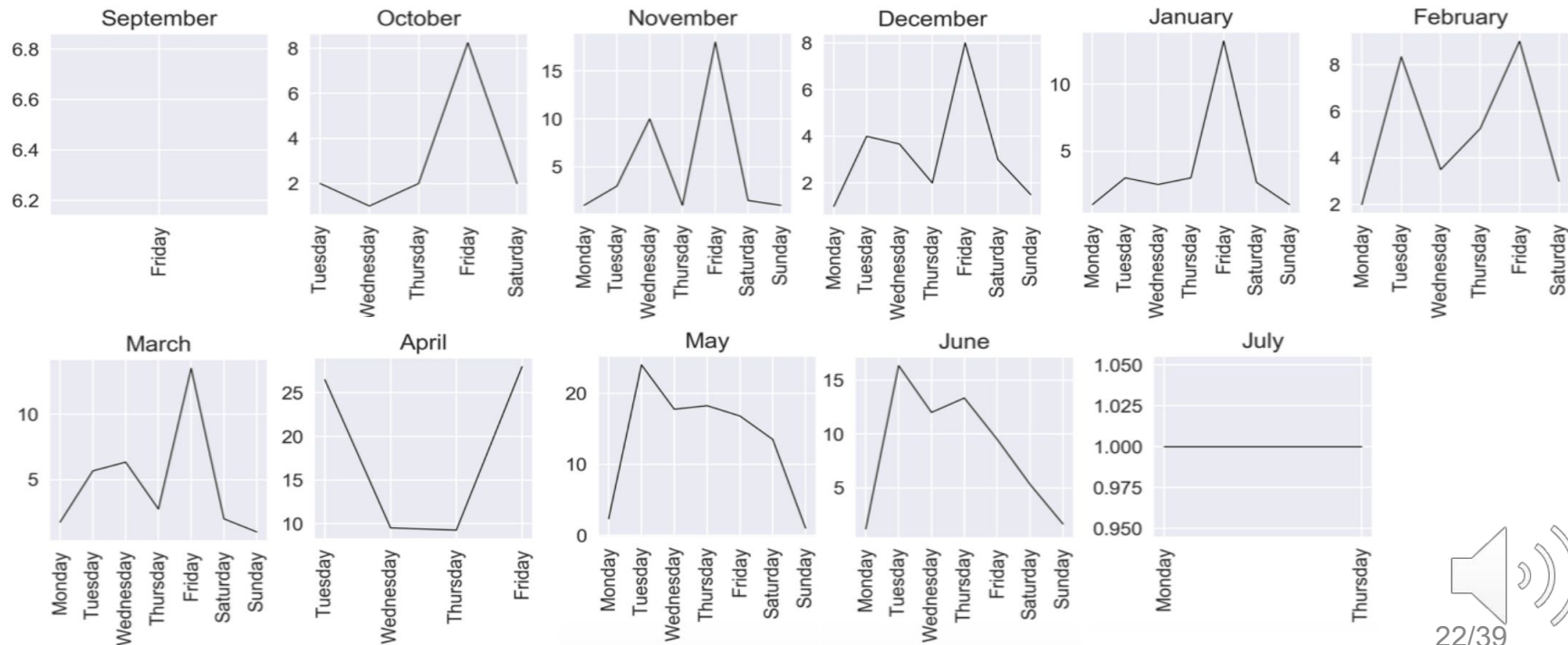|	Low Variance Start/End Times Session	High Variance Start/End Times Session	Small Group Sessions
Count	8,787	**69,554**	5,317
Mean Session Performance	**0.71**	**0.69**	0.67
Mean Class Size	**7.3**	**9.8**	7.7
Mean Class Length (min)	**16.2**	**29.8**	25.4

Gameplay Sessions from Teachers' Perspective per Academic Year

	Students Assisted	Number of Sessions Conducted	Class Length (minutes)	Level Attempts	Time Spent in Game Play (minutes)	Number of Students per Session	Maximum Participants as a Fraction of Total Students in the Session	Class Performance
Mean	~21	~78	28.5	~6	13.82	~9	0.77	0.68

Year-Round Usage of STMath

Average No of Students Playing ST Math on Each Day in the Month

Summary of Observations from Study 1 to Address Research Question 2

1. Most of the time, teachers adopted **High Variance Start/Finish Times Session**. But, **Low Variance Start/Finish Times** had **slightly higher average performance**.

2. Teachers mostly didn't engage all the students in gameplay at the same time.

    Potentially indicative of insufficient technical resources.

3. Gameplay session were **not distributed evenly throughout the year**.

Students played more at the end of the academic year.

Study 2 : Association between session features and session performance
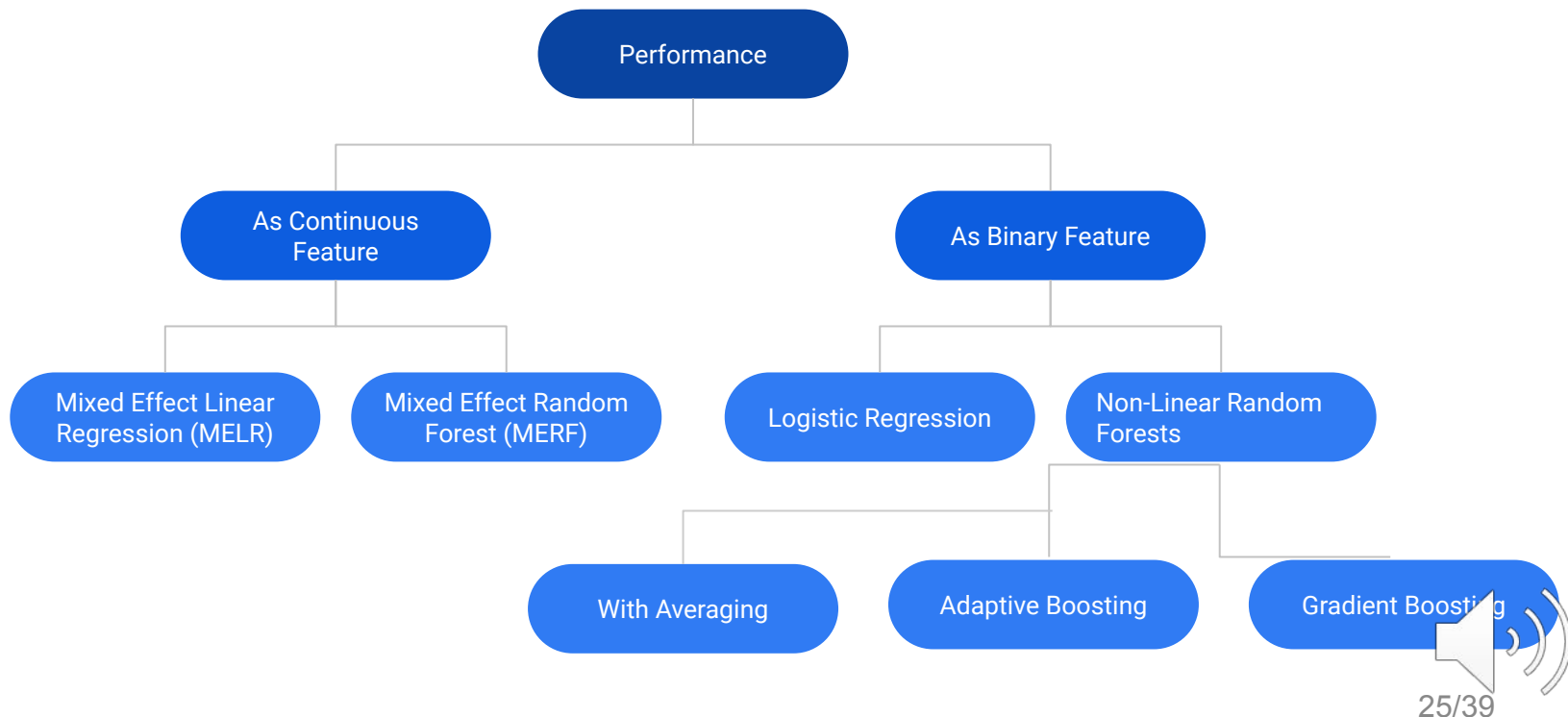
- Built various machine learning models

X = Components of Classroom Integration or the 12 Session Features

Y = Session Performance [Continuous/Binary]

- Goal : Which features have an impact on session performance and how?

Study 2 : Association between session features and session performance

Performance as Continuous Variable - Mixed Effect Regressions

Random Effect Variable (Z) : **grade, teacher id, student id, gameplay content**

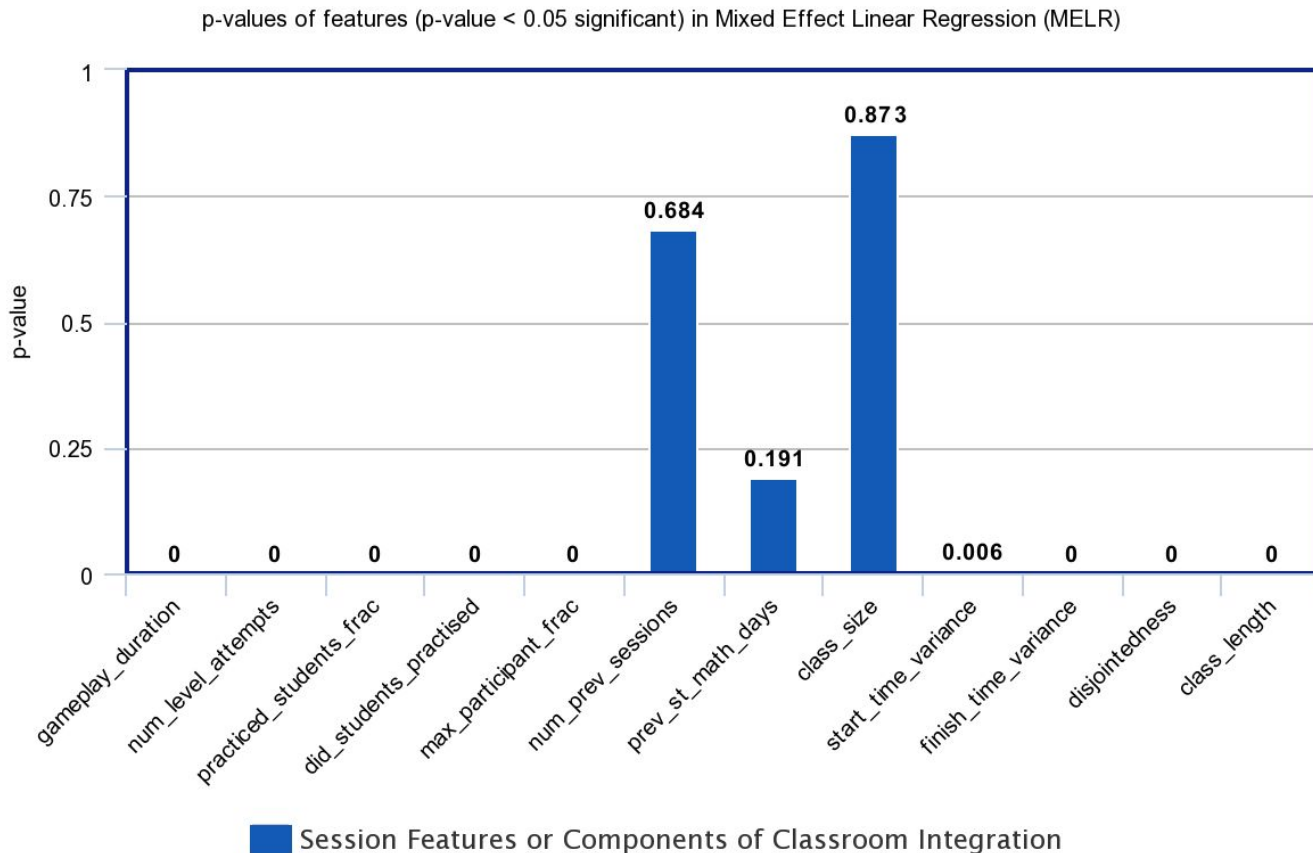- **To eliminate the impact of clusters introduces by these features**

Fixed Effect Variables (X) : **Components of Classroom Integration**

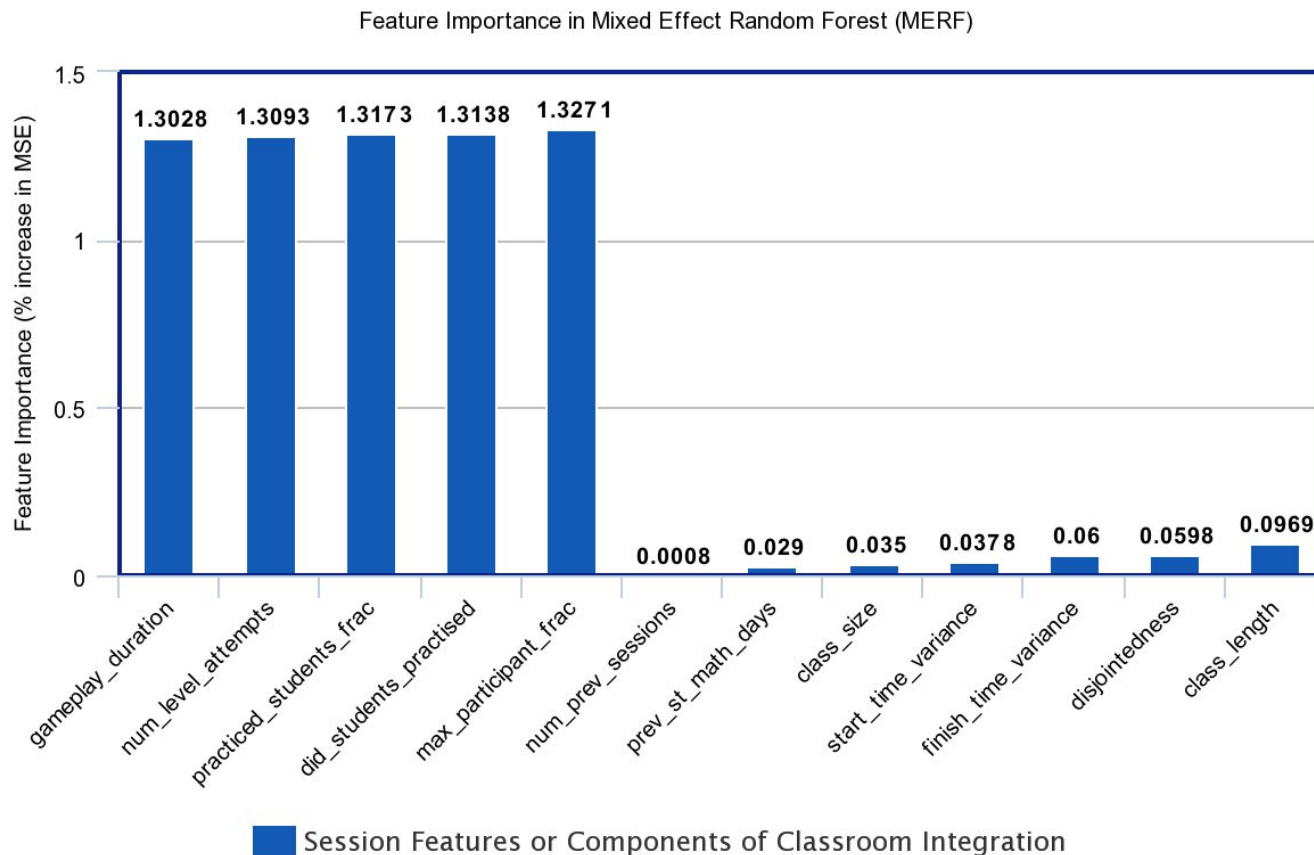Target (Y) : **Session Performance**

Learned Equation : Y = $\color{blue}{f(X)}$ + $b_i Z$ + e

Performance as Continuous Variable (MELR)

p-values of features (p-value < 0.05 significant) in Mixed Effect Linear Regression (MELR)

Session Features or Components of Classroom Integration

Performance as Continuous Variable (MERF)

Feature Importance in Mixed Effect Random Forest (MERF)

Session Performance as Binary Variable

- Session Performance **below 0.7**
  - Marked as **0**

- Session Performance **equals to or above 0.7**
  - Marked as **1**

Binary Predictors Summary

Model	Perf-Rank	Precision	Recall	F1-Score	Accuracy
Logistic Regression	0	0.58	**0.63**	0.60	59%
	1	0.60	**0.55**	0.57	
Random Forest Avg.	0	0.68	0.56	0.61	**65%**
	1	0.62	**0.73**	0.67	
Adaptive Boosting	0	0.64	0.55	0.59	61.86%
	1	0.60	0.68	0.64	
Gradient Boosting	0	0.70	0.52	0.60	**64.8%**
	1	0.61	**0.78**	0.69	

Feature Ranking from Binary Predictors

Ranking of Components of Classroom Integration Obtained from Binary Predictors
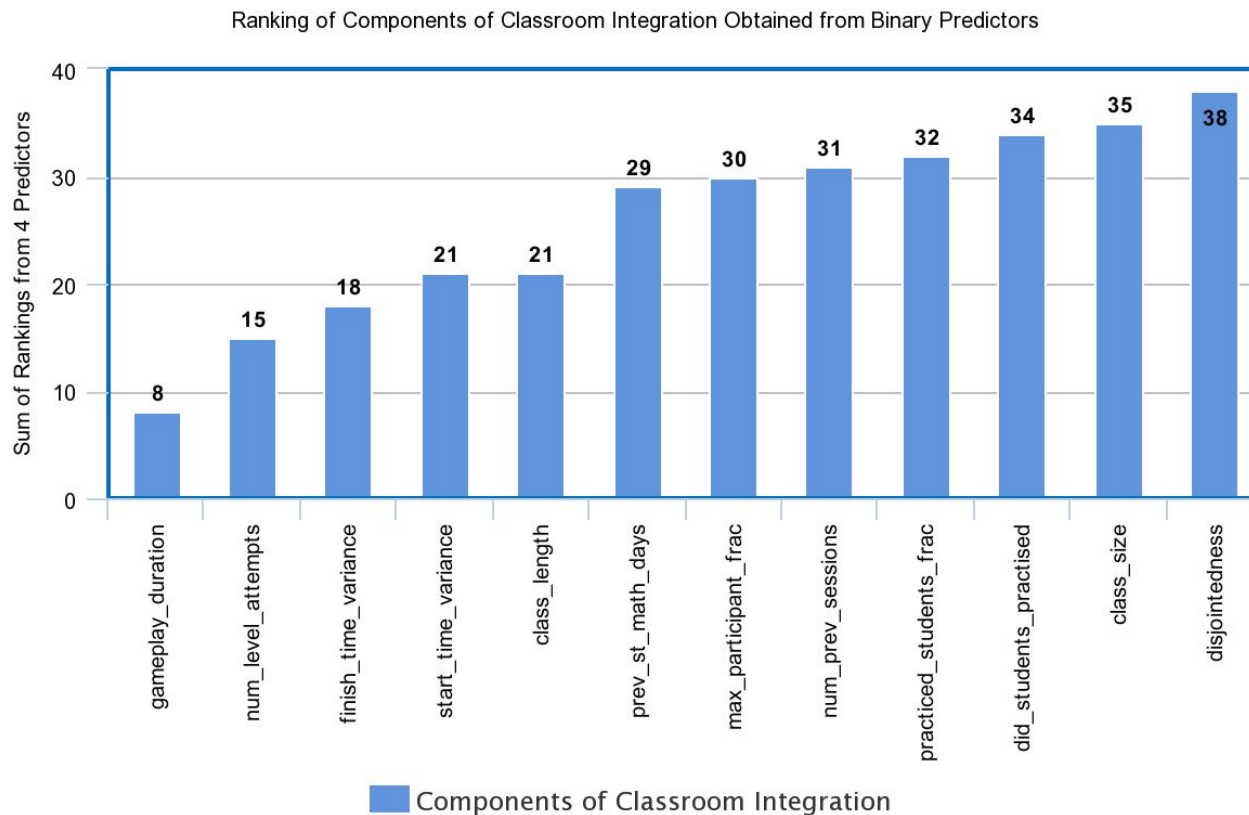
Impact of Gameplay Duration and Number of Level Attempts in Focus

Plot of Performance against gameplay_duration

Plot of Performance against Number of Level Attempt
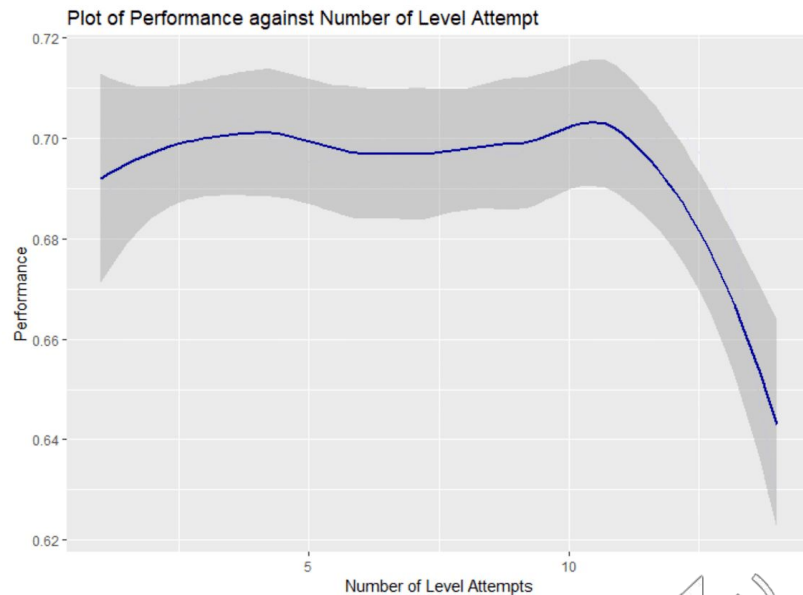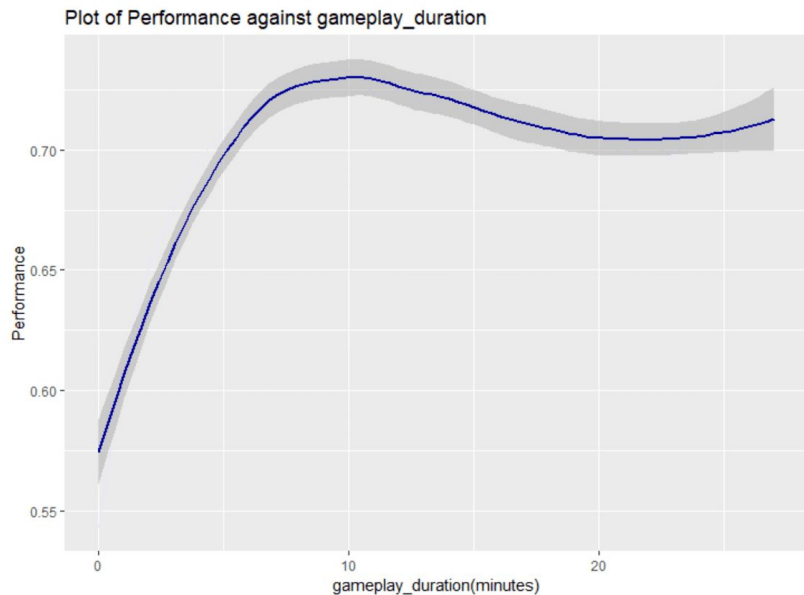
Decision Tree using gameplay_duration and num_level_attempts
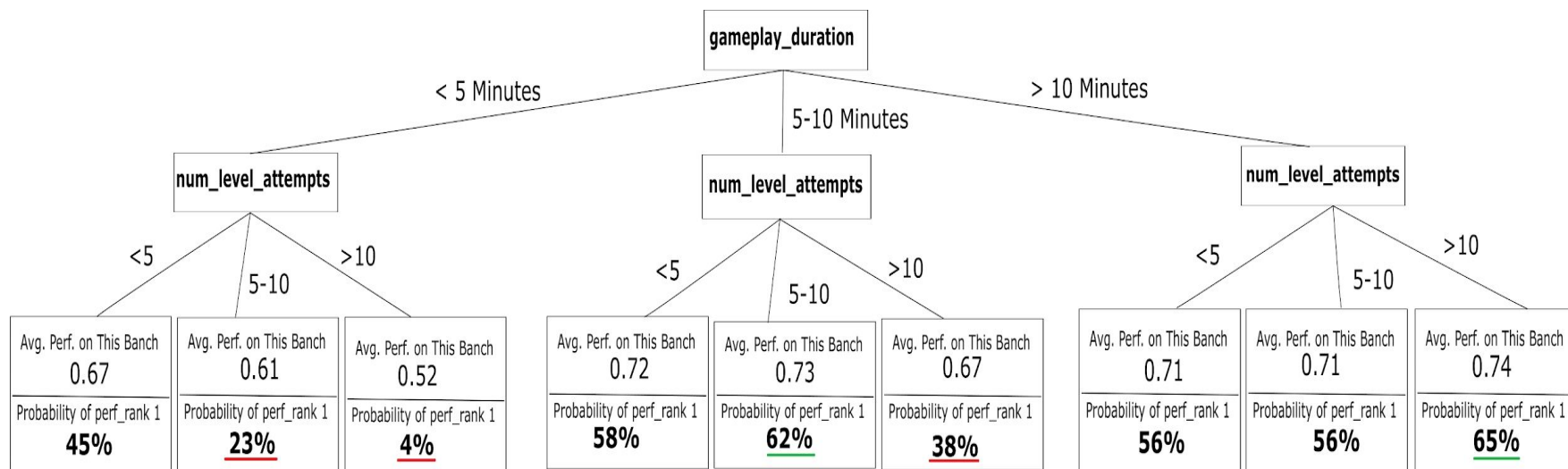
Impact of Components of Classroom Integration on Lowest and Highest Performance in a Session

Feature	Direction of Change	Lowest Performance in a Session	Highest Performance in a Session	Comment
start_time_variance	⬆	⬆	⬇	**Weaker Students Performed Better in High Variance Start/Finish Times Format**
finish_time_variance	⬆	⬆	⬇	
class_size	⬆	⬆	⬇	
num_prev_sessions	⬆	⬆	⬇	**Weaker Students Improved throughout the Semester**
prev_st_math_days	⬆	⬆	⬇	

*** up arrow = increasing ***
*** down arrow = decreasing ***

Summary of Observations from Study 2 to Address Research Question 4

1. All the models agreed only on the association of **num_level_attempts and gameplay_duration**
2. Students performed well when played **moderate number of levels** in a **moderate amount of time**.
3. Lowest performance in a session increases and highest performance decreases with increase in start_time_variance and finish_time_variance
   - **Weaker Students perform better in High Variance Start/Finish times Session**
4. Lowest performance in a session increases as num_prev_sessions and prev_st_math days increases
   - **Weaker students improved gradually throughout the semester**

Overall Summary

1. Students' year-long gameplay activity demonstrated that students play a lot in a few short bursts throughout the year and our findings showed that students do not do well when they play too much in a short time period
   - **Game-play sessions should be evenly distributed throughout the year so that students can focus more on learning rather than just completing the assigned content**

2. Components of Classroom Integration fell short to identify poor performance

Limitations

1. The study is entirely dependent on the session identification strategy
   - No ground truth data on what actually happened in the class
   - Teacher ID is not a perfect variable

2. Explored only ~21% of total sessions

3. No separate analysis for different grade levels

4. Only one timestamp indicating level completion

Future Work

Future work can revise the study to derive more specific suggestions on gameplay session design with -

1. Some **Ground Truth** data about the classrooms
2. Additional data on **Time spent** on specific levels
3. **Accurate Teacher ID**

Thank You for Your Patience!!

Questions???

Email : pshabri@ncsu.edu