Exploring a Flexible Computational Method for Comparing Massive Interaction Data from Science Visualizations

Sweta Karlekar, *Computer Science,* 2017
Faculty Advisor: Dr. Kelly Ryoo, School of Education   Graduate Student Mentors: Emily Toutkoushian & Amanda Swearingen, School of Education

THE UNIVERSITY *of* NORTH CAROLINA *at* CHAPEL HILL

Introduction

Next Generation Science Standards (NGSS) require all students to engage in science practices to deepen their understanding of science (NGSS Lead States 2013). However, **linguistically diverse students**, particularly those from low-income families, often attend public schools that have **limited resources** or limited access to such practices.

Advances in technologies provide tools to automatically capture and analyze students' interaction while engaging in science practices. Such data can be used to create an **adaptive learning environment** with personalized feedback.

- **Current understanding of how students interact with science visualizations** due to their varying knowledge levels is limited, thus making it challenging to create personalized feedback.
- **Current methods of processing interaction data**, such as transition matrices (Mavrikis 2015) and multi-nodal directed graphs (Vovides & Inman 2016), are inflexible and hard to incorporate in newly developed visualizations.

Research Question

How does a new **computational method** capture different **interaction patterns** between students with varying knowledge level?

Methods & Data

This study, which is part of a larger study, involved 80 eighth grade students from two low-income, linguistically-diverse middle schools.

- English Language Learners (ELLs) and non-ELLs in each class were paired to complete the provided inquiry units for 2 weeks.
- Students' interactions with two simulations were automatically logged.

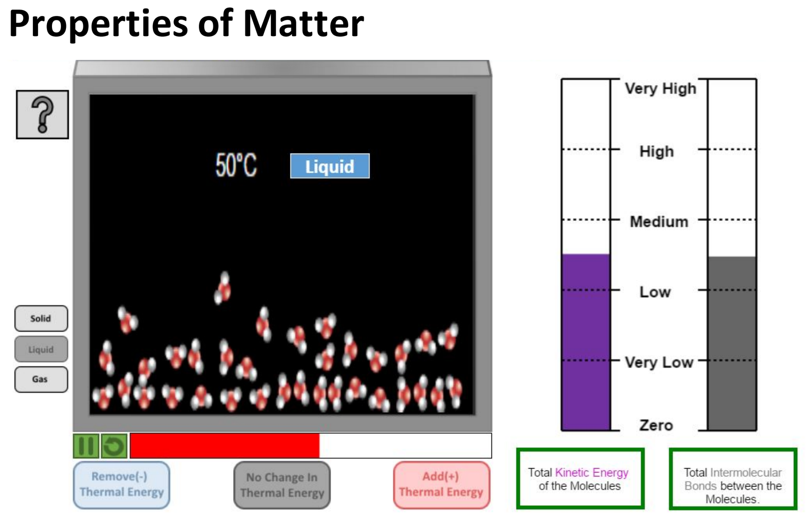
This study explored the differences in interaction data based on students' performance in prediction and reflection activities. In particular, this study focused on the **interaction patterns** captured by a **new computational method**, which uses numerical encoding and Levenshtein edit distance to encode and compare student interaction data.

Study Overview

Student pairs interact with simulations to predict, observe, and reflect → Student's responses & interactions are logged (e.g., button clicks) → Students' prediction and reflection responses are scored → Students' log data are assigned to groups based on response scores → Quantitative analysis of students' interactions → Qualitative analysis of students' interaction patterns

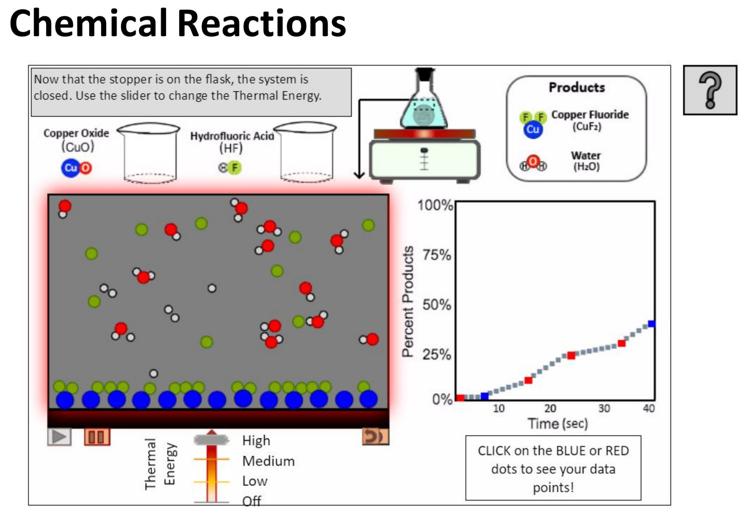
Interactive Simulations

Properties of Matter



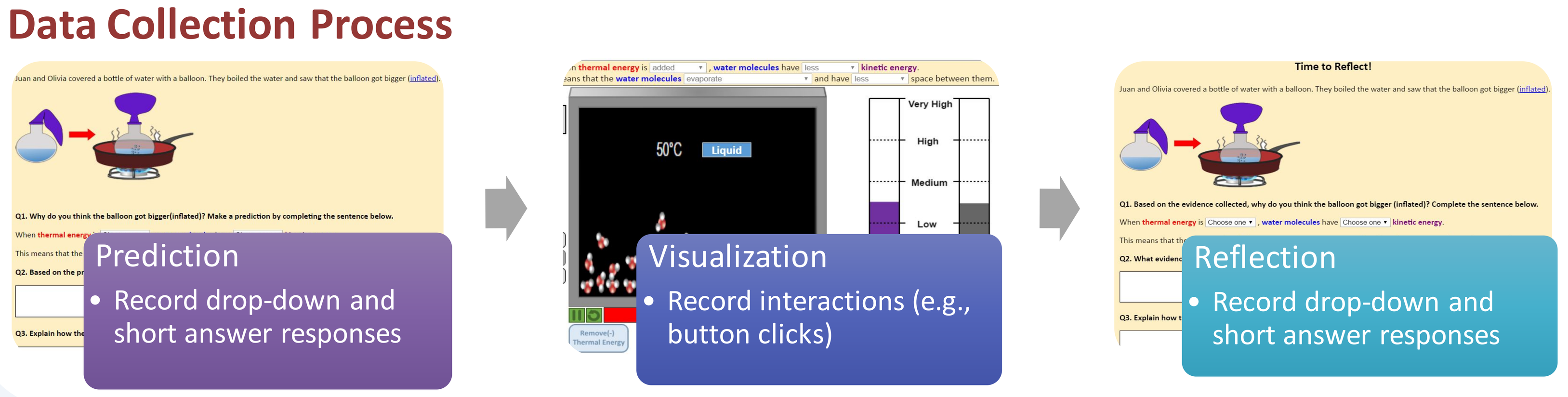
This simulation allows students to visualize the relationships between thermal energy, kinetic energy, molecular movement, and the space between molecules during phase changes.

Chemical Reactions



This simulation allows students to visualize how the relationships between thermal energy, molecular movement, and bond breaking/forming affect the rate of chemical reactions.

Data Collection Process



Data Analysis Procedures & Results

Prediction and Reflection Rubrics

- Prediction and reflection questions were scored independently
- Each prediction and reflection question consisted of three parts: claim, evidence, and reasoning.

Claim — **Total points possible: 1**
- 0 for incorrect claim
- 1 for correct claim

Evidence — **Total points possible: 2**
- 0 for incorrect scientific evidence or off-topic
- 1 for accurate scientific details but no mention of specific evidence
- 2 for accurate and supportive evidence provided in the visualization

Reasoning — **Total points possible: 2**
- 0 for incorrect reasoning or off-topic
- 1 for partial reasoning or does not support evidence provided
- 2 for accurate reasoning that supports the evidence provided

Computational Method

- Step 1: **Group students** based on chosen criteria: prediction & reflection scores
- Step 2: **Assign an** integer value to each interaction
- Step 3: **Convert** interactions list to assigned integer values
- Step 4: **Input** into Levenshtein distance algorithm to find similarity scores

Step 1: Assigning Groups

Prediction Scores → Low Group → Reflection Scores → Low/Low Group, Low/High Group

Prediction Scores → High Group → Reflection Scores → High/Low Group, High/High Group

Steps 2 & 3: Preparing Interaction Data

This **computational method** utilizes the Levenshtein edit distance algorithm's ability to find similarity. This requires:

- Transforming interactions to integers such that the edit distance between each interaction is 1
- Assigning each interaction an integer value
- Converting encoded interactions into strings that contain the list of integers
- Comparing strings using Levenshtein edit distance algorithm

Step 4: Levenshtein Edit Distance

Edit distance is a measure similarity between two strings: Lower edit distances indicate higher similarity. For example:

- It takes 5 edits to transform **INTENTION** into **EXECUTION**
- The edit distance is therefore 5
- The edit distance of this string can be compared to another for similarity

INTENTION (• Delete 'I') → NTENTION (• Substitute 'N' for 'E') → ETENTION (• Substitute 'T' for 'X') → EXENTION (• Insert 'C') → EXECNTION (• Substitute 'N' for 'U') → EXECUTION

Quantitative Data Analysis

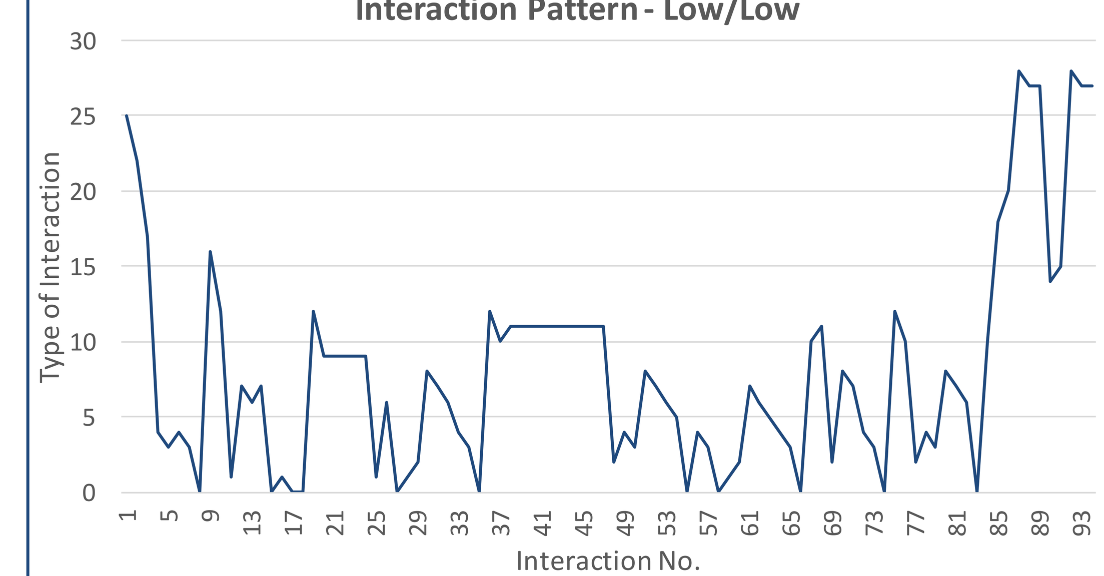
- Students pairs' average Levenshtein distances within each score group was compared
- ANOVA showed statistically significant differences between mean edit distances of the score groups ($p < .05$)
- The results suggest that this **new computational method** shows promise for comparing interaction data as differences in interaction patterns were found between the score groups

Qualitative Analysis of Representative Individuals

- Using the numerical encoding performed previously for the **computational method**, graphs were made to visually compare **interaction patterns**.
- Each interaction is plotted along the time progression of the X-axis against the assigned integer for the interaction on the Y-axis.

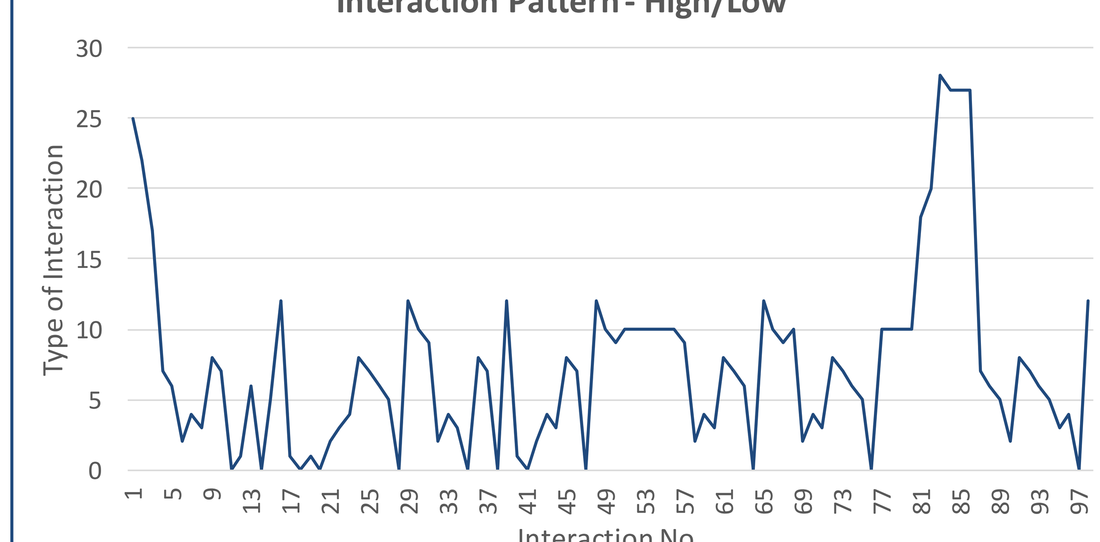
Students with Low Reflection Scores

Prediction Score: 3



Reflection Score: 4

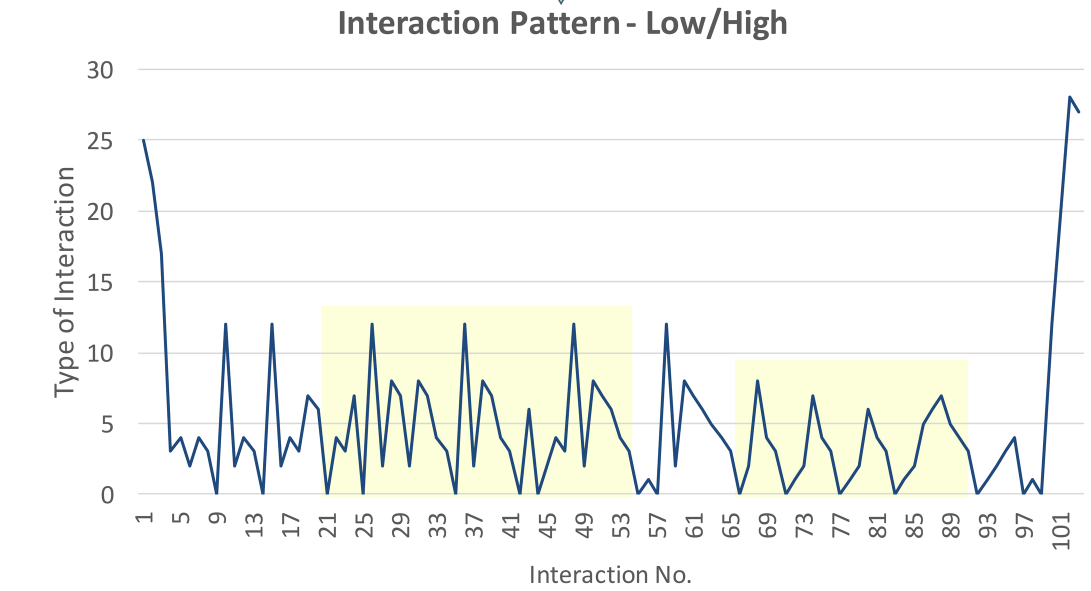
Prediction Score: 6



Reflection Score: 4

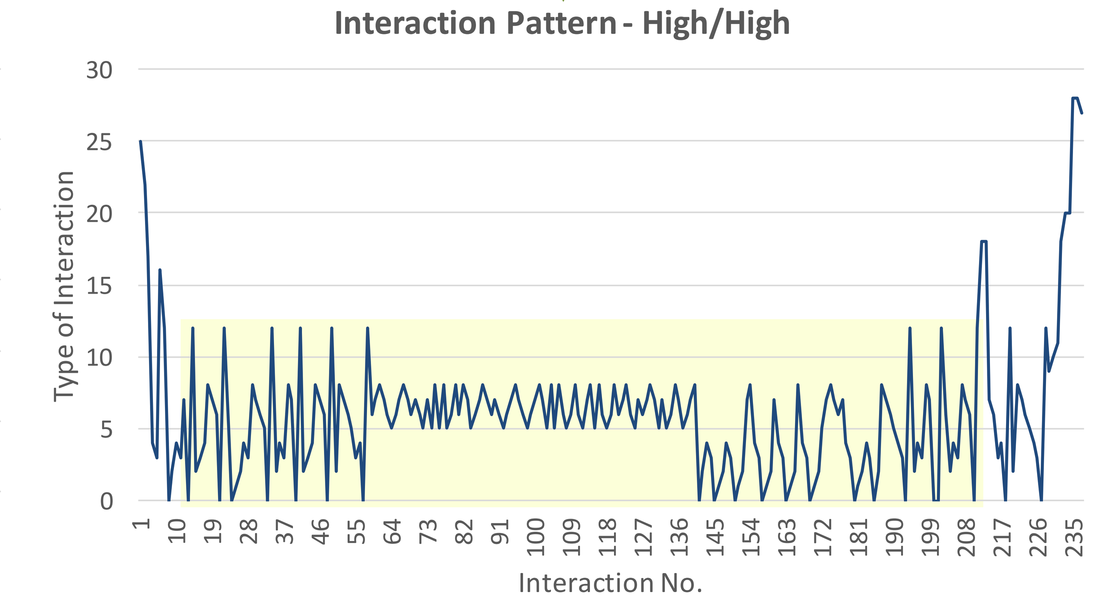
Students with High Reflection Scores

Prediction Score: 3



Reflection Score: 6

Prediction Score: 6



Reflection Score: 7

- Students who scored low on reflection questions often had chaotic interactions that seemingly **random peaks and very little repetition**.
- Students who scored highly on the reflection questions had more **defined, repetitive peaks** in their **interaction patterns**.
- For example, the representative individuals **with illustrated interaction patterns** contain many peaks and troughs, indicating the student viewed the **same information multiple times.**
- However, score groups that scored low on the reflection questions were found to contain dissimilar patterns and had more chaotic **interactions with the visualization.**

Conclusions & Future Research

Conclusions

- The **new computational method** was flexible and easy to implement.
- The interaction data results suggest that students with different levels of knowledge have identifiably **different interaction patterns** while they engage in science practices using dynamic visualizations.
- By better understanding **student interaction patterns** and their ability to predict a high or low reflection score, visualizations can be adapted to guide students towards **productive interaction techniques.**
- **Personalized learning environments** can be automated to provide individual feedback, thus further accommodating each student's learning needs

REFERENCES

Mavrikis, M., Zhu, Z., Gutierrez-Santos, S., & Poulovassilis, A. (2015). Visualization and analysis of students' interaction data in exploratory learning environments. Proceedings of the 24th International Conference on World Wide Web. New York, U.S.: Association for Computing Machinery, pp. 1419-1424. ISBN 9781450334730.

NGSS Lead States. (2013). Next Generation Science Standards: For states, by states. Washington, DC: The National Academies Press. doi:10.17226/18290.

Vovides, Y., & Inman, S. (2016). Elusive learning—Using learning analytics to support reflective sensemaking of ill-structured ethical problems: A learner-managed dashboard solution. *Future Internet.* doi:10.3390/fi8020026

This research was supported by the NSF Research Experiences for Undergraduates (REU) funding.