Proposing Regex Sequence Analysis to Explore Elementary Students Design Practice and Collaboration in Engineering Design Challenges

Mr. Tarun Ambaldhage, University at Buffalo, The State University of New York

Tarun Ambaldhage is a graduate student in Engineering Science and Data Science at the University at Buffalo, The State University of New York. His work focuses on data analytics, machine learning, and sequence-based behavioral analysis in collaborative learning environments, with an emphasis on identifying patterns in student interactions to distinguish productive and unproductive behaviors. He has experience in building data pipelines, conducting exploratory data analysis, and applying clustering and predictive modeling techniques to large datasets.

Dr. Corey T Schimpf, University at Buffalo, The State University of New York

Corey Schimpf is an Assistant Professor and Director of Graduate Studies in the Department of Engineering Education at University at Buffalo. He is a former Division Chair for the Design in Engineering Education Division (DEED) for the American Society of Engineering Education. His research interests include engineering and human-centered design, advancing research methods, and technology innovations to support learning in complex domains. He has a PhD from Purdue University in Engineering Education.

Samuel A Adedeji, University at Buffalo, The State University of New York

Samuel Adedeji is a Ph.D. student in Engineering Education and a Master's student in Computer Science and Engineering at the University at Buffalo, State University of New York. His research focuses on understanding students' affective experiences in engineering learning using fine-grained behavioral data, learning analytics, and machine learning approaches. His work explores how students' actions during engineering design tasks relate to their emotions and learning processes. Samuel holds a bachelor's degree in Chemical Engineering from Ladoke Akintola University of Technology, Nigeria. He also has professional experience as a software engineer. His broader research interests include computing education, educational technology, and the use of data to improve teaching and learning.

Dr. Jutshi Agarwal, University at Buffalo, The State University of New York

Jutshi Agarwal serves as a Research Scientist at the Department of Engineering Education at the University at Buffalo, SUNY. She has a Ph.D. in Engineering Education from the University of Cincinnati with Master's and Bachelor's degree in Aerospace Engineering. Her research is focused on the teaching preparation of graduate students as future engineering faculty, agency of students in open-ended problems, and the professional development of K-12 teachers for teaching engineering design. She strives to identify, adopt, or adapt knowledge from different discipline-based education research (DBER) literature to promote better preparedness and efficacy of educators in engineering at different levels.

©American Society for Engineering Education, 2026

Proposing Regex Sequence Analysis to Explore Elementary Students Design Practice and Collaboration in Engineering Design Challenges

Introduction

This methods full paper seeks to illustrate the utility of adopting regular expressions or "regex" string parsing functions from programming to a type of learning analytics called sequence analysis. Learning analytics is a sub area within the larger field of analytics focused on student's digitally-logged data, such as through learning management systems, which is analyzed to understand and model students learning processes and outcomes [1]. Sequence analysis has been used in the social sciences (e.g., [2]) and education as one approach to learning analytics (e.g., [3]) to uncover meaningful patterns from temporally ordered data, i.e. sequences or subsequences. Sequence analysis is powerful for educational research because these techniques can identify sets of ordered actions students undertake that represent some higher order skill [4] and the distinct ways students may undertake that skill or engage it in productively or unproductively [5]. Sequence analysis is not new – techniques like Markov chains [6] and optimal matching [2] have been used regularly for decades. However, these more established approaches often struggle with identifying complex (e.g., sequences with many different elements) or long sequences. For instance, Markov chains use probability to identify sequences, so when sequences become longer the probabilities of any given sequence become lower and the approach becomes less useful. In this work, we propose a new approach to sequence analysis that utilizes regular expressions (regex) to identify sequences of students learning processes. Regex has the potential to address some of the weaknesses of other sequence analysis approaches and enable the identification of longer and more complex sequences for better student modeling, which can create actionable insights for supporting learners.

Historically, regex has not been used for sequence analysis in learning analytics, therefore this work is adapting regex from its traditional use in computer science for string identification to a data analysis application. To identify sequences, regex uses several analyst-defined operations that specify what matching sequences should or should not include. Sequences that match these criteria are then returned for further inspection or validation. In other words, the analyst can set several rules or criteria by which specific sequences are identified from their data. Traditionally, these features of regex have allowed analysts to automatically locate patterns in large amounts of text, such as email addresses. This makes regex more flexible than other sequence analysis techniques, and potentially more adaptable to students' complex learning processes. In the present work, we aim to provide a proof-of-concept of the utility of regex in learning analytics.

To demonstrate the potential utility of regex for learning analytics, we apply this approach to the context of K-12 engineering education. There has been a growing interest to include engineering as a part of the K–12 curriculum in line with the Next Generation Science Standards [7]. Educational technology is frequently used to support teachers and students to engage in engineering projects in K-12 settings (e.g., [8], [9], [10]). We draw on one such platform here – MindLabs.

MindLabs is an augmented reality educational technology that engages elementary students in science and engineering challenges, to help them develop engineering design and collaborative skills. MindLabs captures student collaborations, design actions, and self-reported emotional states through event data-logging. In a Midwestern school, 223 elementary students collaborated in teams for a learning unit on forces and motion. Students were tasked with designing a system for transporting a ball using MindLabs. Assignments were primarily based on science concepts including forces and motion and culminating with a design challenge. Our analysis draws on their logged actions to perform regex sequence analysis. Our driving question for demonstrating the utility of regex through this K-12 context is:

To what extent can regular expression extract distinct elementary students' engineering design and collaborative practices from learning process logs?

In the remainder of the work, we first review past research on collaborative learning and sequence analysis. Then we describe the context of the study, participants and how we operationalized and used regex for this study, including how data was processed and sequences extracted. Our results highlight students' prevalence of different types of sequences and then discuss the five unique types we identified. The discussion focuses on how the sequences relate to past work on design skills and collaboration, and what advantages and disadvantages regex has compared to other sequence analysis methods, before concluding.

Literature review

Collaborative Learning

Collaborative learning involves students working in groups towards a common goal, by sharing tasks, that leads to individual and collective learning [11], [12], [13]. When students engage in collaborative learning, considering peers as sources of expertise and knowledge [14], it leads to better learning and knowledge retention [15], [16], higher academic achievement [15], [17], and greater critical thinking skills [18]. However, mismatched levels of commitment between team members, social loafing, inadequate communication, and incoherence can lead to dysfunction that poses challenges for both students and teachers [19]. Teachers often have to rely on self and peer assessments to identify dysfunction in team dynamics and evaluate individual contribution. However, these feedback mechanisms are primarily retrospective and can prove unreliable due to challenges in disaggregating various implicit biases of raters and the temporal distance between experience and reflection [20], [21], [22]. Researchers and practitioners recognize the need for real-time feedback mechanisms that can help students in developing collaborative skills in addition to meeting outcomes of learning experiences and are engaged in developing such platforms (e.g., [23]).

Sequence analysis

Sequence analysis is a suite of approaches including Markov chains, n-grams, optimal matching and other approaches that seek to identify or extract a set of temporally ordered elements that represent some phenomena or subject of study [2]. In the social sciences and applied fields like education, the elements in a sequence may represent things like events, actions, or processes. In sequence analysis, there is an assumption that elements of a sequence or sub-sequence (part of a full sequence) are stochastically related to each other (e.g., the presence or absence of some elements is related to or informs other elements) and sequences or sub-sequences reveal something about the activity or people being studied. In the social sciences, sequence analysis has often focused on longer temporal processes, for example, life course trajectories [2] and employment biographies [26]. In contrast, for many studies in education, sequence analysis has been used to study more temporally constrained processes such as problem solving [3], [6], [24] or progress through a course [25].

Application of sequence analysis in educational research has been accelerated by work that argues for creating fine-grained and comprehensive user action or artifact transformation logging systems in digital learning or professional software environments [4], [26]. Platforms that follow these guidelines produce a large amount of temporal data on user-actions or artifact manipulation (e.g., computer-aided design models) that can be fruitfully examined through sequence analysis approaches.

Turning more specifically to sequence analysis approaches, one of the more popular techniques has been Markov chains [6], [24], [27]. Markov chains model the probability of moving from the current state to the next in what is called a transition matrix. These states are the elements of a sequence. While Markov chains can be a useful modeling approach, when there are a large number of unique "states" or many elements exist within a sequence until the next transition (and you need to model all of the combinations), it is difficult to estimate or extract probabilities as they become very small or absent.

Another technique commonly used in the social sciences for sequence analysis is optimal matching [1]. Optimal matching or OM conducts pairwise comparisons of sequences and computes a distance score depending on the number of changes it would take to transform one sequence to another [28]. These changes or operations may involve substituting one element for another, removing an element or adding an element. The technique was originally adapted from work in biology and computer science [29] and has seen some similar applications in education (e.g., see [2]). While innovative, OM has been heavily critiqued (see [2], pp 41-44 for a summary of these critiques). One of the strongest criticisms is that the transformation costs (i.e., removing, deleting or substituting) lack clear guidelines or conceptual grounding on how these should be set. Often the elements in a sequence may be substantially different in quality or qualities (for example, a log of computer-aided design elements may include events like add a roof and analyze the heat transfer properties of a building), rendering a numeric "cost" between elements unclear. Consequently, the transformation costs risk becoming somewhat arbitrary and may either lack sensitivity or be overly sensitive to costs, and in either circumstance they bias what sequences are or are not identified. Thus, although there are several more well-established sequence analysis techniques, many of them suffer from some limitations when dealing with larger sequence lengths or when the types of elements in a dataset vary highly in quality or type.

Methods

Study Context and Participants

This study took place in three fifth-grade classrooms at a suburban elementary school in the Midwestern United States. A total of 223 students took part in the study by engaging in a 7-lesson unit. Most students had little prior experience with engineering design. The lessons were aligned with NGSS physical science standards and focused on forces and motion. The lessons also highlighted basic engineering practices and encouraged students to solve problems through iterative testing and redesign. Teachers had the choice on how to set up student teams, based on their perspective on how their students would best learn. Consequently, team sizes vary from 1-3 in our results. It is also important to note that a small number of students had their teams reassigned, so the total number of students and team sizes do not add up to exactly 223 as above. Over seven one-hour class sessions, students used augmented reality tools to explore ideas related to force and motion. The unit ended with a multi-day design challenge in which teams built a course that moved a ball from a starting point to an ending point using several simple machines.

Description of MindLabs

For this project, students used MindLabs, which is an Augmented Reality (AR) platform where students can collaboratively place tiles (e.g., a pulley or lever) to design systems in an AR space and solve science challenges. Students use a tablet or computer to interact with the platform and data is logged as they interact. MindLabs has a robust learning process data logging system. It captures students' design process and collaborative actions (as they place tiles and build a system together). Additionally, typed entries into the engineering journal, self-reported affective states, and individual-initiated feedback to teammates are also logged. All of these data streams are logged at the device in JSON format and then pushed to a Postgres SQL server where they can be retrieved for post-hoc analysis, viewed by the teacher through an instructional dashboard, or accessed by the system to provide support for students.

Data Analysis

Regular Expression

A regular expression is a sequence of operations that specify a search pattern in text strings used in computer science and programming more generally. It is often referred to as regex in shorthand and is available for most modern programming languages. It is similar to an advanced version of "find" function for text search. Instead of searching text character by character manually, regex allows us to define concise patterns that can detect repetitions, sequences, boundaries, or specific structural rules in large bodies of text or text datasets. Regular expressions rely on a small set of operators to define how patterns are matched. We discuss a few of the key operators here, but for a fuller list of all operators in Python, see [35]. Plus, operator (+) indicates to match or find at least one or more occurrences, and asterisk (*) used to match zero or more occurrences. "^" and "$" matches beginning and end of the string or line, respectively. "(…)" anything between parentheses groups a set of matching characters together.

In this study, regex was used as a method for identifying meaningful behavioral patterns from student interaction logged data generated during assignments across MindLabs, described previously. Using regex, we extracted behavioral patterns such as unequal collaborative contributions (e.g., one student dominating), redo/iteration, and system removal behavior. Student interaction logged data is transformed into structured textual sequences i.e., various important attributes are converted into strings and converted into tokens and to full sequence of actions.

How we used regex

Raw data consists of temporal interaction logs during multiple assignment activities. Each row corresponds to a single interaction event performed by a student within MindLabs. These logs capture fine-grained details about the user, including actions performed, design element involved (e.g., different AR tiles, such as a spring or incline), and relevant contextual metadata as shown in Table 1.

Table 1: Description of attributes from dataset

Attributes	Description
groupId	A team ID with mixture of alphanumeric characters.
studentId	A student ID with mixture of alphanumeric characters.
actionType	Indicates if tile has been added, modified or removed.
tileType	Contains various tile types (spring, lever, …).
tileVariant	Indicates which version of tile is being used.
createdAt	Indicates date & time of action.

To process this data, we first ordered the data temporally to preserve the order of actions and to ensure temporal and structural consistency. Timestamps were converted into a standardized date-time format, enabling reliable chronological ordering of events. Any missing records, invalid, and/or duplicate timestamps caused by known errors in the data logging platform were removed from the dataset to maintain consistency and completeness. Next, the data was encoded from long-strings to single-letters or converted from numerical values to strings datatype which helps in making a full ordered sequence for each team.

- Ex: Lever (tileType) -> as "L" or 123 (studentId) -> as "123"

Event data, however, often contains multiple attributes or metadata which may not easily translate into a single letter or short array of numbers. To address this concern, we developed a token system for transforming a single event data point to a compound token. For these tokens, all the encoded data attributes were joined by an "underscore" to form one single entity. A Token represents one complete action at a particular time and encapsulates most detailed information needed for finding patterns in students' learning behavior. This transformation compresses multiple attributes into a compact and standardized representation.

- Ex: 348_M_L_21_2 -> indicates studentID 348 modified (M) tileType Lever (L) of variant 21 and rotated it clockwise180 (2, where 1 is a directional 90 degs turn).

Tokens for each team are ordered based on their timestamps and are joined using delimiter ':', to form into a full sequence of actions for the respective learning assignments. This sequence representation enabled the analysis to apply regex directly to the sequences in order to detect various behaviors of students interacting while working with fellow peers on assignments.

- Ex: 469_A_S_1_2:469_A_C_0_2:469_M_H_0_o …

Extracted learning process sequences

The team collectively identified five learning process patterns as a starting sample to test regex capabilities and then defined regex operations that could extract sequences that matched the pattern criteria. These patterns reflect learning behaviors identified in past literature or from our experience analyzing and modeling student logged data. We selected five sequences to give a range of tests for this proof of concept. The five specific patterns we selected are listed below.

First, long repetition single-student and second, repetition multi-student are both sequences where the same tile (e.g., an incline plane or a spring) is repeatedly used, either by a single student or multiple students on a team respectively. Third, continual removal involves chains of actions where students/team repeatedly remove tiles (these could be similar or dissimilar tiles) from their design space. Fourth, continual modification is similar to the prior sequence, but involves repeated modification of tiles (for example, changing a specific tile's attributes, like the slope of an incline).

Finally, the redo pattern identifies repetition of same action after certain number of actions. Such patterns could indicate students redo action or iterative design behavior. Table 2 breaks down the regex operators used for the redo sequence search as an illustrative example. We do not list all regex operations due to space limitations.

Table 2: Redo regex operation code

Operations Code	Description
(?:(?<=:)|^)	Start match at sequence beginning or right after a ':' delimiter
([^:]+)	Capture one complete action token, everything until the next ':' delimiter
(?:(?::[^:]+)){{n},}	Match n intervening action-tokens (any tokens, not equal to the first)
(?::|$)	End match at a ':' delimiter or at the end of the sequence

Results

The results are presented in two sections. First, we present an overview through summarized descriptive statistics in Tables 3, 4, and 5 below. Table 3 reports the number of teams (or students working individually, team size = 1) exhibited at least one of a given sequence type. Table 4 reports descriptive statistics (specifically minimum, mean, median, and maximum) for the number of sequences exhibited by teams (size >=1) by a given sequence type. Finally, Table 5 again displays descriptive statistics but for the variations in length or number of actions-tokens within different sequence types by team (size >=1).

Table 3: Number of teams who exhibited at least one of the sequences

Team Size	**Number of Teams**	**Long Repetition Single-Student**	**Long Repetition Multiple-Students**	**Redo**	**Continual Removal**	**Continual Modify**
1	118	41.5%	41.5%	96.6%	37.2%	87.2%
2	62	48.3%	48.3%	91.9%	35.4%	83.8%
3	4	75%	75%	50%	50%	75%

Table 4: Descriptive statistics of frequency count of each sequence type

Team Size	Long Repetition Single Student	Long Repetition Multiple Student	Redo	Continual Removal	Continual Modify
1	[1.0, 1.62, 1.0, 6.0]	[1.0, 1.51, 1.0, 4.0]	[1.0, 1.09, 1.0, 3.0]	[1.0, 1.62, 1.0, 10.0]	[1.0, 1.62, 1.0, 6.0]
2	[1.0, 1.87, 1.0, 9.0]	[1.0, 2.11, 2.0, 5.0]	[1.0, 1.27, 1.0, 3.0]	[1.0, 2.23, 2.0, 17.0]	[1.0, 1.87, 1.0, 9.0]
3	[1.0, 1.33, 1.0, 2.0]	[3.0, 3.0, 3.0, 3.0]	[1.0, 1.0, 1.0, 1.0]	[1.0, 1.67, 2.0, 2.0]	[1.0, 1.33, 1.0, 2.0]

[Minimum, Mean, Median, and Max] Number of Sequences by Type and Team Size

Table 5: Descriptive statistics of sequence length for each sequence type

Team Size	Long Repetition Single-Student	Long Repetition Multiple-Students	Redo	Continual Removal	Continual Modify
1	[5.0, 16.09, 14.0, 61.0]	[7.0, 35.35, 29.0, 194.0]	[6.0, 7.85, 7.0, 24.0]	[5.0, 15.88, 12.0, 100.0]	[5.0, 16.09, 14.0, 61.0]
2	[5.0, 19.74, 14.0, 86.0]	[8.0, 49.87, 44.0, 297.0]	[5.0, 9.5, 8.0, 26.0]	[5.0, 24.96, 17.0, 210.0]	[5.0, 19.69, 14.0, 86.0]
3	[6.0, 26.67, 11.0, 63.0]	[57.0, 89.5, 89.5, 122.0]	[8.0, 8.5, 8.5, 9.0]	[7.0, 33.0, 29.0, 63.0]	[6.0, 26.67, 11.0, 63.0]

[Minimum, Mean, Median, and Max] Length of Sequences by Type and Team Size

Next, we discuss each sequence type in more depth, referring to Tables 3-5 as well as some additional analysis. We present the visualizations for a few of the sequence types (redo and continual modification) to give a window into what some specific sequences look like. We are only able to provide a few visualization examples due to space limitations.

Long Repetition & Long Repetition with multiple students sequences:

Long Repetition is when the same action-token occurs recurrently in a full sequence. As each action-token encodes studentID identifier along with other action attributes, consecutive repetition of an identical token typically indicates repeated actions performed by same student without interruption. Similarly, Long repetition with multiple students' sequence is similar to long repetition sequence but allows for multiple studentID, this type of sequence indicates consecutive repeated actions as well. Both these sequences have almost the same statistics as shown in Table 4 and 5 (Columns 2 and 3). This indicates that most repetitions were potentially made by one student.

Redo Sequence:

Redo sequence captures interaction patterns in which an earlier action is repeated after some intervening actions. In redo sequence, returning to same action type or operation after performing other actions in between indicating a revisitation of prior decisions. This type of sequence could be made by one or more students.

Almost 90% of teams exhibit redo behavior, suggesting revisiting earlier actions is an important aspect of the design process. Average length of redo sequences is nearly double that of other sequences, around 35 action-tokens for single student teams. Figure 1 below shows a visualization depicting some actions for a redo sequence. At the beginning of the sequence, student 347 added a screw, then a target-archery tile in next action-token. Similarly, some tiles were added or modified, and screw was removed mid-sequence; later on, the screw was added after going through multiple changes, showing "redo" of same action again at the end of sequence.

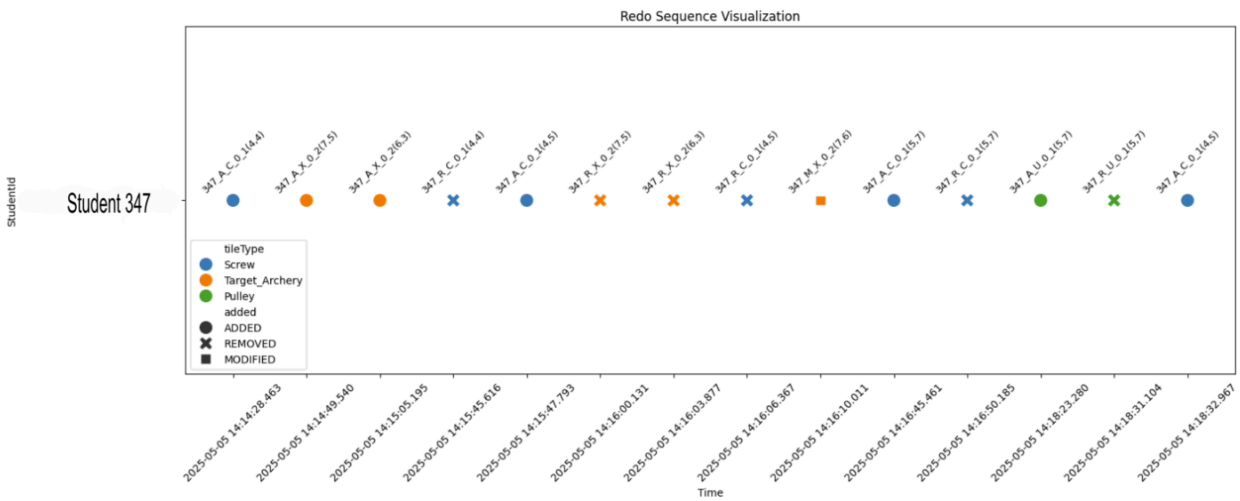


Figure 1: A Sample Redo Sequence over time

Continual Removal & Modify Sequences:

Continual removal sequence captures periods where teams perform multiple deletion actions back to back with little or no switching to other operations. This type of behavior depicts that the team is actively undoing prior work, eliminating parts of the systems, cleaning up initial built design or reacting to mistakes. Continuous modify sequence captures teams repeatedly performing modification actions consecutively such as adjusting, rotating or altering components. This pattern typically means a team is undergoing iterative refinement of current design independent of which member performs actions.

103 teams have on average 2 continual modify sequences with average length of 15 action-tokens for teams with a single student, whereas average number of action-tokens of continual removal sequences is almost half the average length of the modify sequences. Figure 2 below shows action-tokens for one continual modification; the values in parentheses are x and y

coordinates of tile placed in the grid plane. It shows student 363 working on modifying the tileType by using different versions of lever. Student 371 was working on the Ball Anchor by placing it at different coordinates in grid space, after student 363's work is complete.

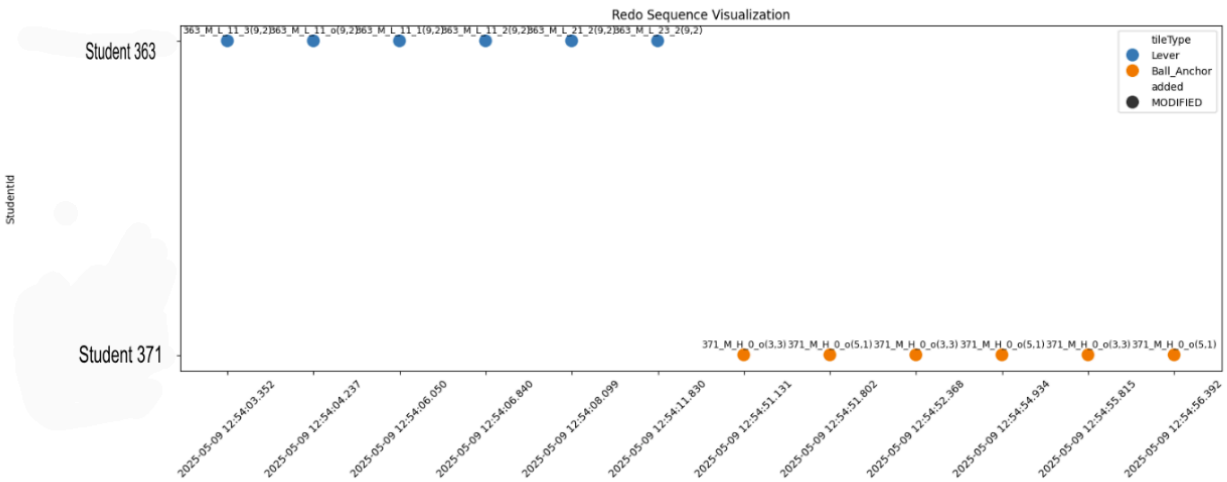


Figure 2: A Sample Continual Modify sequence over time

Discussion and Conclusion

The goal of this study is to examine whether regular expressions can be used to identify how students learn while working on science and engineering design tasks. Below we provide some preliminary interpretations of what these sequences may indicate about student learning or difficulties they are encountering.

Long repetition sequences occur when students had interacted with the system without meaningful changes to their design. This could be student making adjustment to a lever and doing this without major changes and many times. This type of behavior could suggest that students are not making any progress and are confused with the task. *Redo sequences* occurred when students returned to an earlier component after working with other tiles. For example, students might add a spring, test the system, and then go back to adjusting a wheel-and-axle or pulley they had used before the spring. This pattern relates to iterations in students design process, an important skill given the complexity of design problems [30] It may also indicate students were learning through experimentation and revision. These sequences may also indicate that students were using feedback from the system to rethink earlier decisions. This behavior is central to engineering design and shows that the sequences may capture iterative and/or reflective learning processes. *Continual removal sequences* occurred when students removed multiple components, such as removing several planes, springs, or pulleys in sequence. These actions may show recognition of systematic errors or challenges in their design or may also reflect students' frustration (e.g., [19]) with the task or their design, showing students are trying to restart their design. *Continual modification sequences* occurred when students made repeated changes to existing components, such as slightly moving a screw, rotating a lever, or adjusting the length or position of a pendulum. Instead of starting over, students made small adjustments to improve their designs. This behavior may reflect persistence and small-scale modifications or optimization (e.g., see [31] of their existing system. It may also represent small trial and error changes to try to incrementally improve the system.

The action sequences identified in this analysis and discussed above begin to reveal some ways in which students were engaged or potentially disengaged in the learning while working on the science and engineering challenge. Our language above remains tentative as sequences may represent different design practices or have different influencing factors. For instance, the continual removal could represent students recognizing a large issue with their system and restarting or it could represent student frustration with the task. A key insight from this preliminary regex extraction is that regex searches may need to incorporate more sequence context (e.g., more actions before or after the sequence) to better unveil what kind of practice it represents.

More broadly, the differences in sequence types across teams align with prior research showing that teams approach the same design problem in different ways [32], [33]. The range of sequence types identified also demonstrates the flexibility of regular expressions in capturing varied forms of student design behavior. While this analysis focused on relatively simple sequence patterns, the results suggest that this approach can be extended to more complex sequence discovery and modeling of student learning processes. This work also responds to calls in prior research to better use data generated by educational technologies to understand how learners engage with tasks [30, 31].

From a methodological perspective, regular expressions offer several advantages over more traditional sequence analysis approaches such as Markov chains and optimal matching. Because this approach relies on analyst-defined criteria rather than probabilities or frequent transitions, it can identify meaningful but less common sequences. This is especially useful for design tasks, where student behaviors can vary widely [5], [34] and often unfold over longer action sequences. Compared to optimal matching, regular expressions allow analysts to focus on specific action patterns without requiring the estimation of transformation costs [1]. However, this approach also has limitations. Defining meaningful regular expressions requires time, careful thought, and a strong understanding of how student learning behaviors appear in action data. For simpler patterns, other methods may be faster or easier to apply. As a result, regular expressions should be viewed as a complementary tool rather than a replacement for all sequence analysis methods.

More generally, this approach has the potential to support teaching and learning in engineering design. The findings show that regular expressions can effectively extract clear and interpretable action sequences that reflect how elementary students learn engineering design through experimentation, reflection, persistence, and revision while working with simple machines. Like all studies, this work has limitations. The analysis relies on software-generated action logs and does not capture verbal discussion among students or interactions with the teacher. Future work will extend this approach by incorporating additional data sources, such as affective states, team feedback, or learning outcomes. Further research will also explore whether specific action sequences are associated with productive or unproductive learning outcomes, using assessments, final artifacts, or predictive modeling approaches. Moreover, work remains to compare the approach proposed here to more established sequence analysis techniques like Markov chains or optimal matching. Finally, this method can be applied to other learning technologies to better understand student skills and practices and to provide more targeted support for learners and educators.

References

[1] G. Siemens, "Learning Analytics: The Emergence of a Discipline," *American Behavioral Scientist*, vol. 57, no. 10, Art. no. 10, Oct. 2013, doi: 10.1177/0002764213498851.

[2] B. Cornwell, *Social Sequence Analysis: Methods and Applications*. New York, NY: Cambridge University Press, 2015.

[3] E. Kleinman *et al.*, "Analyzing Students' Problem-Solving Sequences: A Human-in-the-Loop Approach," *Learning Analytics*, vol. 9, no. 2, pp. 138–160, Aug. 2022, doi: 10.18608/jla.2022.7465.

[4] C. Schimpf and M. H. Goldstein, "Large data for design research: An educational technology framework for studying design activity using a big data approach," *Front. Manuf. Technol.*, vol. 2, p. 971410, Oct. 2022, doi: 10.3389/fmtec.2022.971410.

[5] C. Vieira, M. Hathaway Goldstein, Ş. Purzer, and A. J. Magana, "Using Learning Analytics to Characterize Student Experimentation Strategies in Engineering Design," *Journal of Learning Analytics*, vol. 3, no. 3, Art. no. 3, 2016, doi: 10.18608/jla.2016.33.14.

[6] C. McComb, J. Cagan, and K. Kotovsky, "Capturing Human Sequence-Learning Abilities in Configuration Design Tasks Through Markov Chains," *Journal of Mechanical Design*, vol. 139, no. 9, Art. no. 9, Jul. 2017, doi: 10.1115/1.4037185.

[7] NGSS Lead States, "The Next Generation Science Standards: For States, By States," The National Academies Press, Washington D.C., 2013.

[8] C. T. Schimpf *et al.*, "Board 394: Supporting Secondary Students' Engineering Front-End Design Skills with the Mobile Design Studio," in *American Society of Engineering Education Conference Proceedings*, Portland, OR, 2024.

[9] S. Sung, X. Ding, R. Jiang, E. Sereiviene, D. Bulseco, and C. Xie, "Using artificial intelligence teaching assistants to guide students in solar energy engineering design," *Journal of Geoscience Education*, vol. 72, no. 4, pp. 347–366, Oct. 2024, doi: 10.1080/10899995.2024.2384340.

[10] J. Chiu, A. Gonczi, X. Fu, and M. D. Burghardt, "Supporting Informed Engineering Design across Formal and Informal Contexts with WISEngineering," *International Journal of Engineering Education*, vol. 33, no. 1, 2017.

[11] E. F. Barkley, C. H. Major, and K. P. Cross, *Collaborative learning techniques: A handbook for college faculty*. Josey-Bass, 2014.

[12] J. Clark, "Stimulating collaboration and discussion in online learning environments," *The Internet and Higher Education*, vol. 4, no. 2, pp. 119–124, Apr. 2001, doi: 10.1016/S1096-7516(01)00054-9.

[13] P. Dillenbourg, "What do you mean by 'collaborative learning'?," in *Collaborative-learning: Cognitive and computational approaches*, P. Dillenbourg, Ed., Elsevier, 1999.

[14] M. Henry, "Developing critical, creative and collaborative skills," in *Learning in the Digital Age*, 2nd Edition., Routledge, 2014.

[15] S. Hussain, S. Anwar, and M. I. Majoka, "EFFECT OF PEER GROUP ACTIVITY-BASED LEARNING ON STUDENTS' ACADEMIC ACHIEVEMENT IN PHYSICS AT SECONDARY LEVEL," *International Journal of Academic Research*, vol. 3, no. 1, 2011.

[16] P. T. Terenzini, A. F. Cabrera, C. L. Colbeck, J. M. Parente, and S. A. Bjorklund, "Collaborative Learning vs. Lecture/Discussion: Students' Reported Learning Gains,"

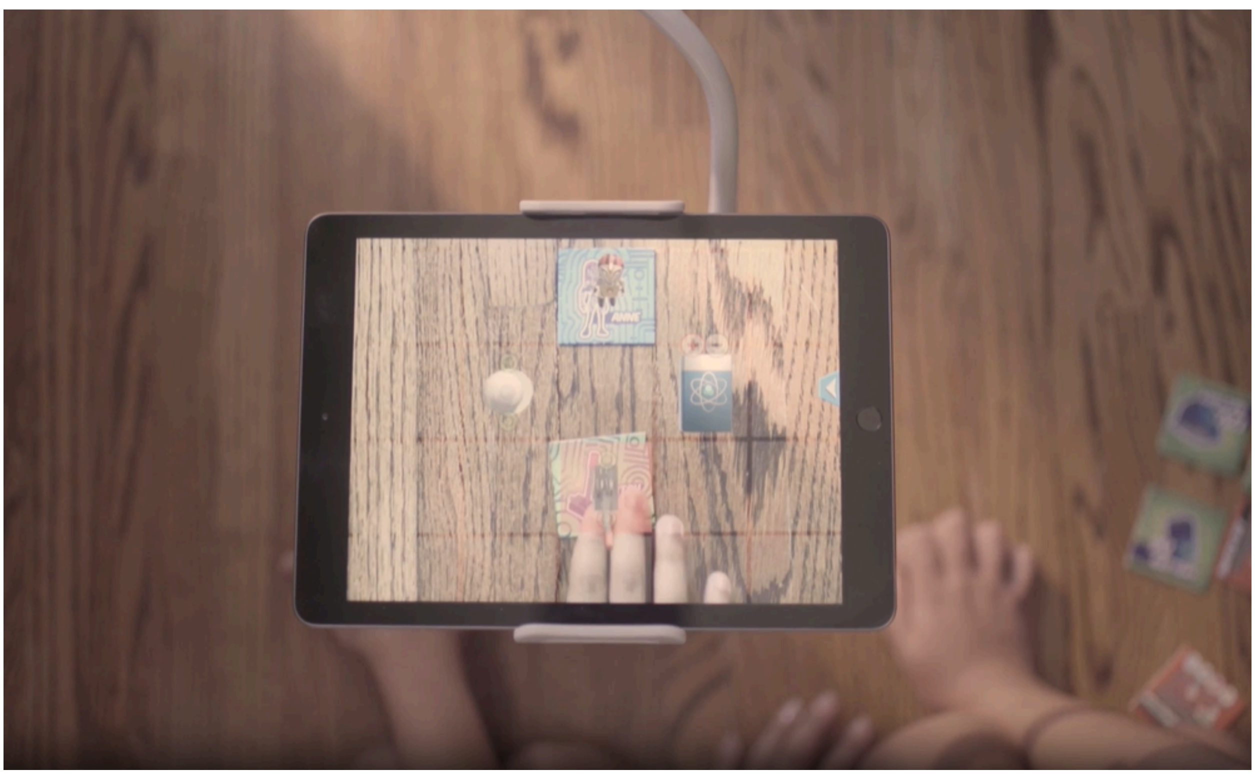
Journal of Engineering Education, vol. 90, no. 1, pp. 123–130, 2001, doi: 10.1002/j.2168-9830.2001.tb00579.x.
[17] L. Springer, M. E. Stanne, and S. S. Donovan, “Effects of Small-Group Learning on Undergraduates in Science, Mathematics, Engineering, and Technology: A Meta-Analysis,” *Review of Educational Research*, vol. 69, no. 1, pp. 21–51, Mar. 1999, doi: 10.3102/00346543069001021.
[18] A. Gokhale, “Collaborative learning enhances critical thinking,” *Journal of Technology education*, vol. 7, no. 1, 1995.
[19] C. Schimpf, J. Agarwal, N. Sunar, L. Smith, A. Thompson, and S. A. Mehdi, “WIP: Leveraging Learning Analytics to Explore Elementary Students Collaboration and Affect in Engineering Design Challenges,” in *2024 IEEE Frontiers in Education Conference (FIE)*, Washington, DC, USA: IEEE, Oct. 2024, pp. 1–5. doi: 10.1109/FIE61694.2024.10892823.
[20] D. Davis *et al.*, “Assessing Team Member Citizenship in Capstone Engineering Design Courses,” *International Journal of Engineering Education*, vol. 26, no. 4, pp. 771–783, 2010.
[21] T. O’Neill *et al.*, “Introducing a scalable peer feedback system for learning teams,” *Assessment & Evaluation in Higher Education*, vol. 44, no. 6, pp. 848–862, 2019, doi: 10.1080/02602938.2018.1526256.
[22] J. H. Stonewall, “Development and evaluation of training to increase student perceptions of fairness in peer assessment,” Doctoral Dissertation, Iowa State University, 2022. Accessed: Jan. 21, 2026. [Online]. Available: https://dr.lib.iastate.edu/handle/20.500.12876/0zEyJ15z
[23] T. Breideband *et al.*, “The Community Builder (CoBi): Helping Students to Develop Better Small Group Collaborative Learning Skills,” in *Computer Supported Cooperative Work and Social Computing*, Minneapolis MN USA: ACM, Oct. 2023, pp. 376–380. doi: 10.1145/3584931.3607498.
[24] C. Schimpf, M. Goldstein, R. Adams, J. Chao, S. Purzer, and C. Xie, “Work in Progress: A Markov Chain Method for Modeling Student Behaviors,” in *2018 ASEE Annual Conference & Exposition Proceedings*, Salt Lake City, Utah: ASEE Conferences, Jun. 2018, p. 31262. doi: 10.18260/1-2--31262.
[25] R. Hlavatý and L. Dömeová, “Students’ Progress throughout Examination Process as a Markov Chain,” *IES*, vol. 7, no. 12, p. p20, Nov. 2014, doi: 10.5539/ies.v7n12p20.
[26] A.-P. Vrolijk and A. Olechowski, “Computer-aided design as a research instrument for engineering design: supporting multiple methods by observing artifacts, actors, and actions,” *Journal of Engineering Design*, vol. 36, no. 12, pp. 2151–2182, Dec. 2025, doi: 10.1080/09544828.2024.2432835.
[27] M. H. Rahman, M. Gashler, C. Xie, and Z. Sha, “Automatic Clustering of Sequential Design Behaviors,” in *Volume 1B: 38th Computers and Information in Engineering Conference*, Quebec City, Quebec, Canada: ASME, Aug. 2018, p. V01BT02A041. doi: 10.1115/DETC2018-86300.
[28] C. Brzinsky-Fay and U. Kohler, “New Developments in Sequence Analysis,” *Sociological Methods & Research*, vol. 38, no. 3, pp. 359–364, Feb. 2010, doi: 10.1177/0049124110363371.
[29] A. Abbott and J. Forrest, “Optimal Matching Methods for Historical Sequences,” *Journal of Interdisciplinary History*, vol. 16, no. 3, p. 471, 1986, doi: 10.2307/204500.
[30] C. Schimpf and C. Xie, “Characterizing Students’ Micro-Iterations Strategies through Data-Logged Design Actions,” in *2017 ASEE Annual Conference & Exposition*

Proceedings, Columbus, Ohio: ASEE Conferences, Jun. 2017, p. 28027. doi: 10.18260/1-2--28027.
[31] J. L. Chiu, J. P. Bywater, T. Karabiyik, A. Magana, C. Schimpf, and Y. Y. Seah, "Comparing Optimization Practices Across Engineering Learning Contexts Using Process Data," *J Sci Educ Technol*, vol. 33, no. 1, pp. 143–155, Feb. 2024, doi: 10.1007/s10956-023-10080-x.
[32] C. J. Atman, "Design timelines: Concrete and sticky representations of design process expertise," *Design Studies*, vol. 65, pp. 125–151, Nov. 2019, doi: 10.1016/j.destud.2019.10.004.
[33] C. J. Atman, R. S. Adams, M. E. Cardella, J. Turns, S. Mosborg, and J. Saleem, "Engineering Design Processes: A Comparison of Students and Expert Practitioners," *Journal of Engineering Education*, vol. 96, no. 4, Art. no. 4, Oct. 2007, doi: 10.1002/j.2168-9830.2007.tb00945.x.
[34] C. Schimpf, R. Castellani, and M. H. Goldstein, "Uncovering pre-college students reflection strategies for solving complex engineering design problems," *Int J Technol Des Educ*, vol. 35, no. 3, pp. 907–928, Jul. 2025, doi: 10.1007/s10798-024-09930-5.

Acknowledgments

This work is funded by NIH project number 2R44GM134813-02. All opinions are those of the authors and do not reflect the views of NIH.

Appendix MindLabs in Use



Appendix Figure – Students using MindLabs with a tablet