Jiang, B. et al. (Eds.) (2025). Proceedings of the 33rd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

How do students design? A data-driven approach for understanding students' engineering design behaviors

Shruti JAIN*, Namrata SRIVASTAVA, Clayton COHN, Naveeduddin MOHAMMED & Gautam BISWAS
Vanderbilt University, USA
**shruti.jain@vanderbilt.edu*

Abstract: Engineering design is a natural addition to STEM+C education, as it motivates students to integrate science and computing while engaging in problem-solving tasks. However, K-12 students face challenges in applying reasoning and critical thinking skills to find "optimal" design solutions. In this paper, we investigate students' behaviors when they are engaged in engineering design tasks by analyzing the sequences of the design solutions they generate to find an "optimal" solution. We collected data from sixth-grade students (N=88) working in pairs in an NGSS-aligned earth sciences curriculum. Using Markov Models, we visualized students' solution progressions and identified instances where they became stuck in sub-optimal states. Our results show that (1) students' scores improved significantly from pre-test to post-test; (2) students' problem-solving behaviors could be categorized as *design space explorers* and *conservative designers*; and (3) students' problem-solving behaviors influenced the quality of their final designs. Additionally, our findings underscore the importance of supporting design space exploration as a key aspect of scaffolding engineering design activities.

Keywords: STEM+C, Engineering design, NGSS, Learning Analytics

1. Introduction

Engineering design (ED) helps advance STEM+C education, enabling students to combine scientific and computational concepts and practices to solve real-world problems. ED builds 21st-century skills, such as critical thinking, creativity, collaboration, and communication (Ali & Tse, 2023). The National Research Council (2012) and the Next Generation Science Standards (NGSS Lead States, 2013) underscore the importance of integrating ED into K–12 science curricula to foster problem-solving skills, design thinking, and deeper engagement with core scientific practices.

Science-integrated engineering curricula offer a promising approach to helping students learn by applying science concepts to meaningful design tasks (Kelly & Cunningham, 2019). Through computational modeling and simulation-based learning environments, students can apply their disciplinary knowledge in authentic contexts and develop strategies to search the design space and find an optimal design solution that meets specified constraints (Zhang et al., 2019). However, students often find engineering design tasks challenging. They must understand the scientific principles and also navigate complex design spaces where trade-offs, such as cost versus performance, must be carefully considered (Purzer et al., 2015). Prior research shows that students struggle to identify and manage these trade-offs and often rely on trial-and-error approaches instead of systematic reasoning or science-based justifications (Purzer et al., 2015; Zhang et al., 2019; Hutchins et al., 2021). Compounding this challenge, teachers often lack formal training in engineering design, making it difficult for them to support their students (Radolff & Capobianco, 2021). As a result, students often find it difficult to develop and apply systematic principles and critical thinking skills that support generating high-quality or optimal engineering design solutions.

In this paper, we examine the design solutions created by sixth-grade students participating in the Water Runoff Challenge (WRC), a three-week STEM+C curriculum that seamlessly combines earth science, engineering, and computational thinking (Basu et al., 2022). Students investigate the water absorption and runoff characteristics of different surface materials after heavy rainfall. They learn through hands-on experiments and then develop conceptual and computational models that estimate water runoff for various materials and rainfall amounts, using their absorption capacities and the principle of conservation of matter.

The students work together in pairs to extend their computational models to redesign a schoolyard, aiming to reduce water runoff while meeting constraints related to cost, accessibility, and functionality for the play areas. They are encouraged to explore the design solution space and submit their "best" designs that satisfy all specified criteria. By analyzing the sequence of design solutions students generate as they strive for their "best" solutions, our research seeks to understand how students navigate the solution space, their awareness of the design constraints, and how their problem-solving patterns affect the quality of their final designs. Specifically, we analyze data collected from the WRC learning environment to address the following research questions:

RQ1: How effective is the WRC curriculum in helping students develop engineering design knowledge? To answer RQ1, we analyzed students' learning gains derived from their pre- and post-test scores.

RQ2: How do students' search patterns for design solutions vary as they work toward the optimal solution? To address RQ2, we (i) applied an Expectation Maximization Clustering algorithm to characterize their search patterns, and (ii) studied their problem-solving behaviors using Markov models.

RQ3: How are the different problem-solving behaviors related to the quality of the design solution reported by the students? To address RQ3, we developed scoring schemes to compare the final design solutions of students from the two different clusters.

The results demonstrate that the WRC curriculum effectively supports student learning in science, computational thinking, and engineering design. Findings highlight the need for scaffolding design space exploration and trade-off analysis to help students optimize their solutions. Insights from this work can inform curriculum design, and the development of personalized scaffolds to foster effective integration of engineering design into K–12 classrooms.

2. Literature Review

NGSS advocates for integrating engineering practices in K-12 education (NGSS Lead States, 2013); research on students' engineering design (ED) processes primarily focuses on high school students. For instance, Yu et al. (2020) used structural equation modeling to show that the design process mediates the link between scientific knowledge, critical thinking, and design quality. Purzer et al. (2015) studied high school students designing energy-efficient buildings, highlighting challenges in trade-off analysis. Xing et al. (2023) applied learning analytics to identify challenges related to behavioral features in an "energy-plus" model that was used to design a house with minimal energy requirements.

The limited research on how middle schoolers engage with ED includes Bowen et al. (2016), who noted that while prior knowledge affects initial performance in virtual bridge simulations, iterative practice can help fill these knowledge gaps. Similarly, Du et al. (2025) have studied the dynamics between science concepts and design behaviors of different performance groups using Markov Chain analysis. Montgomery et al. (2020) used click-stream data to calculate sub-behaviors– exploration and systematicity– for understanding students' optimization behaviors. Still, these cumulative metrics fail to capture temporal transitions from exploratory to systematic optimization. Prior work by Zhang et al. (2019) used 3-D visualizations of individual students' trajectories to demonstrate how middle school students struggle with navigating trade-offs. While trying to satisfy one constraint, they end up compromising on another. Zhou et al. (2021) observed that high-performing middle school students effectively integrated all verbal, visual, and physical modalities in a toy design workshop. However, such analysis needs to be done individually and is not scalable (e.g., Kelly & Cunningham, 2019;

Wendell et al., 2017). We need to look at groups of students as a whole to find generalizable trends in students' problem-solving behaviors. To our knowledge, the process of evolution of students' designs in a simulation-based design environment remains underexplored.

Evaluating and providing feedback on ED solutions is challenging, as the ED activities comprise several objective and subjective design criteria. Goldstein et al. (2016) developed a scoring criterion for ED projects, taking the subjectivity into account by adopting a Trade-off Value approach that compares designs based on their performance on technical, economic, and human factors. However, the suggested approach uses percentile calculation, which can only evaluate an ED solution relative to other students' solutions. On the other hand, the scoring criteria in previous works (Zhang et al., 2019) did not fully consider human factors, such as accessibility, of the ED challenge in the WRC curriculum. Hence, we adopted the two frameworks to develop an objective metric to evaluate the ED solutions while accounting for technical, economic, and human factors. It is important to understand how middle school students' design behaviors evolve over time, and how we can analyze these behaviors on a scale. In this research, we aim to extend our understanding of middle school students' problem-solving behaviors in ED activity by studying the progression of students' design solutions.

3. The WRC Curriculum

The WRC curriculum (Basu et al., 2022) includes 12 lessons over three weeks, with daily 45-minute classes conducted in-person, combining traditional pen-and-paper tasks with activities in a computer-based learning environment. Students start with a pre-test and finish with a post-test to evaluate their learning gains. The curriculum covers earth science concepts such as the conservation of mass, rainfall, absorption, and runoff, helping students understand the challenges of urban runoff. They investigate how different surfaces manage rainwater and engage in an Engineering Design Challenge to design a schoolyard that reduces runoff after a rainfall of 2 inches. Students create designs and develop models to simulate and improve water runoff solutions based on the given criteria on the ED interface, as shown in Figure 1.

Students design a low-runoff schoolyard within a \$750,000 budget in an interactive computer environment. The schoolyard is visualized as a 4×4 grid with 12 editable squares selectable from six materials: concrete, permeable concrete, natural grass, artificial turf, wood chips, and poured rubber. Each material varies in cost, absorption, and accessibility. Poured rubber is expensive (cost=\$187,500), but it offers high absorption (1.2 inches) and is fully accessible. In contrast, wood chips (cost=\$37,500) provide moderate absorption (1.0 inches) but lack accessibility. To solve the design challenge, students need to navigate these trade-offs. As students adjust their designs, the cost updates in real time. They can evaluate their design by clicking the "Test Design" button and entering a rainfall amount to visualize total absorption and runoff values. The Design History table (see Figure 1) tracks all tested designs. The engineering design challenge has 6^{12} potential solutions, and approximately 6^{6} meet the runoff, cost, and block constraints detailed in Table 1.

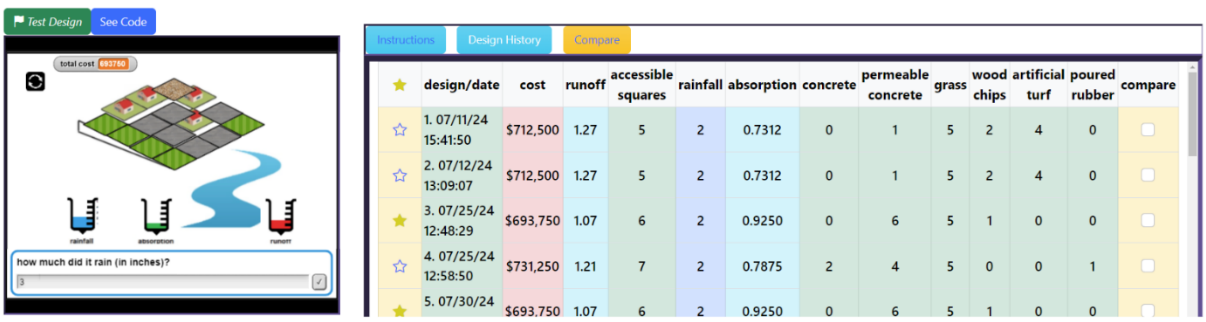

Figure 1. WRC Engineering Design (ED) Environment. The design on the left represents the optimal solution containing 6 permeable concrete, 5 natural grass, and 1 woodchips square.

4. Methodology

This section presents the study design, data collection, measures, and analytical methods, including scoring pre- and post-tests to calculate learning gains, qualitative states, and quantitative scores for analyzing students' engineering design solutions, and the statistical and clustering methods used for understanding design behaviors.

4.1 Study Design and Participants

Our study involved 170 sixth-grade students from an urban public school in the southeastern U.S., comprising 63% White, 17% Black/African American, 7% Hispanic, 9% Asian, and 4% other races, with 51% female and 49% male students. The Vanderbilt University Institutional Review Board approved the study with parental and student consent. The two participating science teachers underwent professional development before the study and co-taught the WRC curriculum with researcher support. The curriculum featured individual and collaborative tasks where students worked in pairs. For engineering design activities, pairs used a shared user ID in the web environment and submitted their final designs in the shared notebook. Students could freely create and test as many solutions as they wished during three class periods. The pairs generated 2 to 47 design solutions (*M=19.86, SD=10.99*). We analyzed pre-post tests, submitted designs, and design sequences from the design history logs collected in the WRC environment. After excluding incomplete records, our dataset comprises 88 students' data (44 pairs) and 874 generated engineering design solutions.

4.2 Measures

Learning gains: To answer RQ1, we analyzed students' responses to validated pre- and post-test included in the WRC curriculum. The tests measure students' knowledge in Science, Computational Thinking (CT), and Engineering Design (ED), and the tests have a section on each topic (Basu et al., 2022). Responses were graded by a researcher and an external grader using validated rubrics. After two rounds, the graders achieved an average Cohen's Quadratic Weighted Kappa (QWK) of 0.9, indicating strong inter-rater reliability (McHugh, 2012).

States of Engineering Design (ED) Solutions: To answer RQ2 and assess the quality of students' ED solutions, we organized their designs hierarchically by runoff levels and compliance with the design constraints. Designs were categorized based on runoff generated from two inches of rainfall: those below the median runoff value across all design solutions were classified as Low Runoff, while others were labeled as High Runoff. Next, we assessed the designs by determining their compliance with the design constraints. Designs that met neither the cost criterion nor the blocks criteria were labeled as Nothing Satisfied. Those that met one of the two criteria were identified as Cost Satisfied or Blocks Satisfied. Designs fulfilling both criteria were classified as All Satisfied. This process resulted in 8 categories, but we combined the high runoff designs into one state due to their smaller number.

In summary, we classified the designs into five states: (i) *High Runoff* (above-median runoff); (ii) *NS-Low Runoff* (low runoff, no design constraints met); (iii) *CostS-Low Runoff* (low runoff, only budget criterion met); (iv) *BlockS-Low Runoff* (low runoff, meeting blocks criteria only); and (v) *AllS-Low Runoff*, for designs meeting all constraints, deemed high quality. This categorization provided insights into students' search patterns, including how they navigate the design space, prioritize constraints, and approach design challenges. While these states primarily reflect solution quality in terms of constraint satisfaction and do not directly measure cognitive processes, analyzing transitions between them using Markov Models allows us to infer consistent problem-solving patterns and strategies.

Final Design Solutions Score: To answer RQ3, we developed a scoring scheme to evaluate students' final designs systematically. Following Goldstein et al (2016), we calculated sub-scores for the design challenge's technical (minimizing runoff under two inches of rainfall), economic (keeping redesign costs below $750,000), and human factors (ensuring accessibility, grassy field, play area, and parking).

Figure 1 shows the optimal design for this challenge, achieving the lowest possible runoff of 1.07 inches with a cost of $693,750 while meeting the *accessibility* and *facilities criteria*, which is collectively termed the *blocks criteria*. Our 12-point scoring system assigns up to 5 points for the runoff value, with full points for the optimal runoff and 0 for incomplete solutions. The *cost criterion* contributes up to 3 points: 2 points for staying under budget, plus one for improving cost-efficiency towards the optimal design cost. The *blocks criteria* are scored binarily: 1 point for meeting the *accessibility criterion* (at least six accessible squares) and 1 point each for satisfying the grassy field, play area, and parking/hardcourt requirements. This scoring scheme, detailed in Table 1, was used to assess the final schoolyard designs submitted by the students at the end of the Engineering Design activity. As it was developed post-hoc, students relied on their judgment to evaluate their design solutions during the activity.

Table 1. *Design constraints of the ED task and the proposed scoring scheme.*

Criteria	Criteria and Score Description	Sub-score
Runoff	Design should minimize runoff for 2 inches of rainfall. Scored on a linear scale (0-5), with the optimum runoff value getting a score of 5, and higher runoff solutions getting lower scores.	5
Cost Criterion	Design must not exceed the budget of $750,000. 2 points for meeting the budget constraint, and a linear scale (continuous, 0-1) for getting closer to the optimal cost.	3
Accessibility Criterion	Design must have 6 squares made of accessible materials (concrete, permeable concrete, artificial turf, or poured rubber). Binary score for meeting the accessibility criterion.	1
Facilities Criteria	**(a) Field**: Design must have 4 squares of grassy field made of grass or artificial turf.	1
	(b) Play area: The Design must have 1 square for the play area made of wood chips or poured rubber.	1
	(c) Parking and Hardcourt: Design must have 5 squares for parking and hardcourt made of concrete or permeable concrete. Awarded binary scores for meeting each of the Facilities Criteria.	1
	Total Score	12

4.3 Data Analysis Methods

4.3.1 Effectiveness of the WRC Curriculum - Statistical Tests

To evaluate the effectiveness of the curriculum, we assessed whether the scores for each section (Science, CT, and ED) on the pre-post tests followed a normal distribution using the Shapiro-Wilk test. Because the data was not normally distributed, we used the non-parametric Wilcoxon signed-rank test to compare student performance on pre- and post-tests.

4.3.2 Identification of Problem-Solving behaviors - FOMM and Clustering analysis

We identified clusters based on problem-solving behaviors using a two-step method involving a First-Order Markov Model (FOMM) and Expectation-Maximization (EM) clustering, following a methodology inspired by prior process mining research (Gatta et al, 2017; Srivastava et al, 2022). First, we classified students' designs into the five states, representing the quality of design solutions (detailed in Section 4.2). We then calculated the sequences of design states for each student pair – the sequences represent the progression of quality of the designs generated iteratively over multiple sessions. Next, using the pMineR package in R (Gatta et al, 2017), we trained a First-Order Markov Model (FOMM) to estimate the transition probabilities between the states, creating transition matrices for each pair's sequence. These transition matrices were input to an EM clustering algorithm to identify distinct problem-solving profiles. The Elbow method suggested an optimal solution of k=2 clusters. Model convergence was monitored using the log-likelihood change across iterations, and cluster performance was evaluated using within-cluster and between-cluster distance metrics derived from the FOMM-based representations. To further interpret the identified clusters, we employed descriptive statistical analysis to summarize key characteristics and differences among the clusters, and FOMM process mining to investigate the transitions between different design states within each cluster. The transition matrices and subsequent clustering reveal consistent patterns in design changes, suggesting intentional problem-solving rather than random modifications.

4.3.3 Influence of Design Behaviors on Final Design Solutions - Statistical Tests

We used descriptive analysis and Mann-Whitney U tests to investigate how different design behaviors influenced the final design solutions that students submitted in their notebooks.

5. Results

5.1 RQ1: Effectiveness of the WRC curriculum

The pre- and post-test analysis for the 88 students showed that the intervention produced significant learning gains ($p<0.001$) for all three domains− Science, CT, and ED. The accumulated average scores (*maxscore=40*) improved from *25.47 (SD=5.95)* to *30.34 (SD=4.64)* points, showing a significant improvement *(W=264.5, r=−0.751, p<0.001)*. Science scores *(maxscore=12)* improved from *7.76 (SD=2.59)* to *8.83 (SD=2.33) (W=599.5, r=−0.603, p=0.001);* CT scores (*maxscore=13*) improved from *8.06 (SD=2.96)* to *10.55 (SD=2.02) (W=268.5, r=−0.749, p<0.001)*; and ED scores *(maxscore=15)* from *9.64 (SD=2.46)* to *10.97 (SD=2.14) (W=612.0, r=−0.597, p<0.001).*

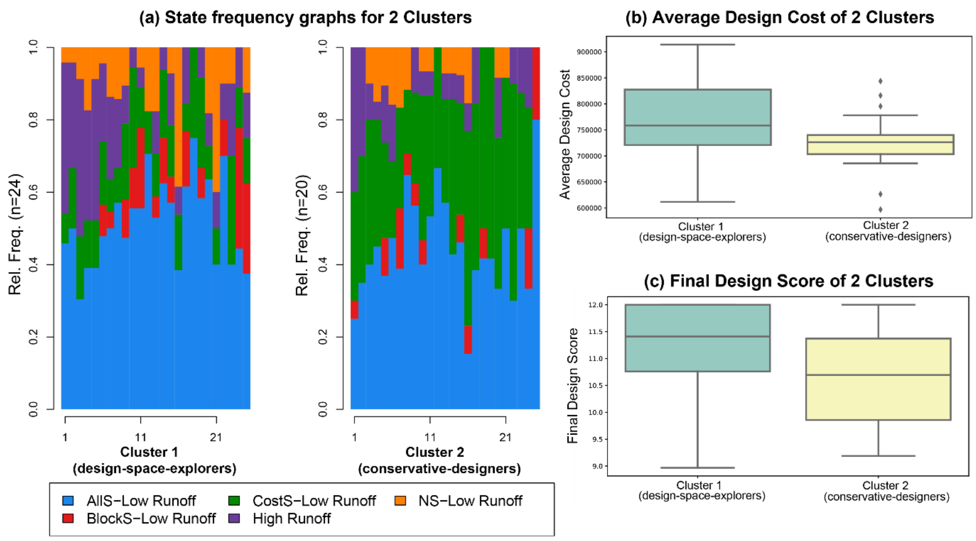

Figure 2. (a) State frequencies over design iterations for the two clusters. Box plots of (b) average cost of generated designs; and (c) Final Design Scores of students in the two clusters.

5.2 RQ2 (i): Identifying Problem-Solving Behaviors

Using the clustering approach (see Section 4.3.2), we identified two distinct clusters of student pairs characterized by their exploration strategies. Figure 2(a) shows state frequencies over design iterations for the two clusters. Notably, pairs in Cluster 2 spent more time in the *CostS-Low Runoff* state, indicating a stronger emphasis on cost optimization. The box plot (Figure 2(b)) also shows that the average cost of designs in Cluster 2 was lower (*M=$726,023.45, SD=55,513.41*) than Cluster 1 (*M=$776,305.04, SD=73,957.01*). This difference was statistically significant (*p=0.02, Mann–Whitney U=339.0, Z=2.333, r=0.352*). Given their cost-focused and cautious strategy, we refer to Cluster 2 as *conservative designers*. In contrast, students in Cluster 1 explored a wider range of design states, demonstrating a more exploratory approach; we therefore refer to this cluster as *design space explorers*.

5.3 RQ2 (ii): Understanding Students' Engineering Design Process

To analyze the exploration strategies of the two clusters—*conservative designers* (*N=20*) and *design space explorers* (*N=24*)—we constructed First-Order Markov models (FOMM) based on the sequences of their design solutions. Figure 3 presents a comparative FOMM for the two clusters, where states (see Section 4.2) represent the quality of design solutions, while edges indicate transitions between states with corresponding probabilities of transition.

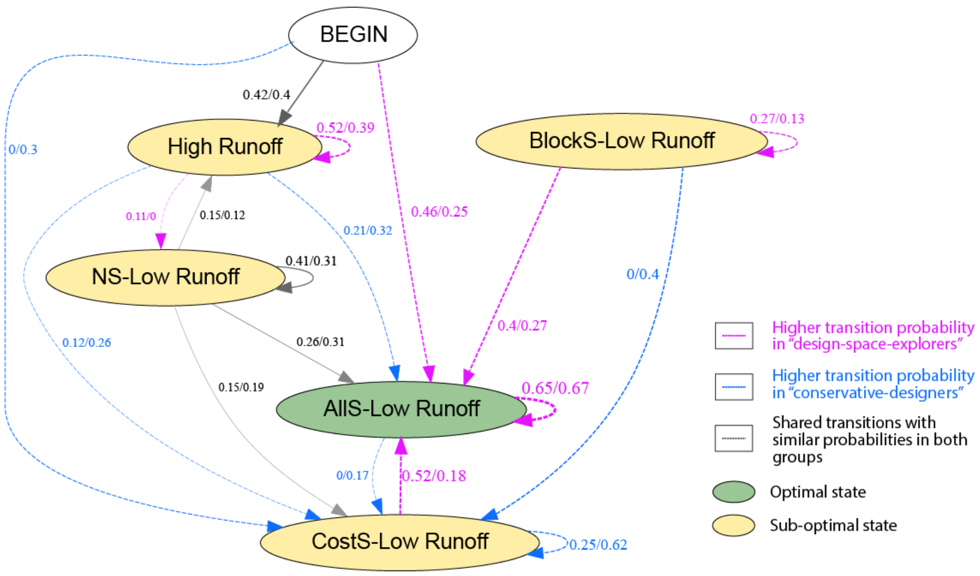

Figure 3. First-order Markov models for "*design space explorers*" and "*conservative designers*". Only transitions with probabilities > 0.1 are shown. Labels show the transition probability for *design space explorers* / *conservative designers*. Colored edges indicate probability differences >0.1 (pink: higher for *design space explorers*; blue: higher for conservative designers).

The *conservative designers* typically began from suboptimal states, such as *High Runoff* (*prob=0.4*), and only a few student pairs achieved the optimal state (*AllS-Low Runoff*) early in their search process (*prob=0.25* of generating the optimal design in the first step). The students initially generated a sequence of *High Runoff* solutions, often not succeeding in improving their solutions (*prob=0.39*) or moving to another sub-optimal state before eventually finding an optimal solution. Interestingly, there are no incoming edges to the *NS-Low Runoff state*, suggesting students achieved low runoff through various approaches and did not backtrack once they had satisfied at least one of the design constraints.

A recurring theme among the *conservative designers* is their focus on cost-driven transitions. The *CostS-Low Runoff* state is the only sub-optimal state with incoming edges from every other state, indicating that students consistently prioritized optimizing the cost of their designs regardless of their starting point. Moreover, the only outgoing edge from the *CostS-Low Runoff* state leads to the optimal state, suggesting that students made a concerted effort to improve cost by either experimenting within the *CostS-Low Runoff* state (*prob=0.62*) or they transitioned to the optimal state (*prob=0.18*) by meeting the *blocks criteria* as well. The back and forth between the optimal *AllS-Low Runoff* state and the suboptimal *CostS-Low Runoff* state shows their cost optimization behavior.

On the other hand, *design space explorers* began their exploration with either optimal(*prob=0.46*) or *High Runoff* states (*prob=0.42*). Once in the *AllS-Low Runoff* or *High Runoff* states, *design space explorers* were likely to remain there (with probabilities of *0.65* and *0.52*, respectively). The *design space explorers* were more willing to experiment with different designs early in their process, as evidenced by the loop between the *NS-Low Runoff* and *High Runoff* states, suggesting that they were more willing to sacrifice low runoff in their pursuit of satisfying the design constraints. A commonality in both the *conservative designers* and the *design space explorers* is that there are no incoming edges to *the BlockS-Low Runoff state,* indicating that both groups prioritized the other design constraints, like the *cost criterion* over the *blocks criteria*. The Markov model in Figure 3 also suggests that *design space explorers* prefer to minimize runoff while addressing design constraints simultaneously, for instance, in the *NS-Low Runoff* state, where *design space explorers* had a strong tendency to remain (*prob=0.41*) before eventually transitioning directly to the optimal state (*prob=0.26*).

5.4 RQ3: Influence of design behaviors on Final Design Solutions

We graded the final designs submitted in their notebooks using the scoring criteria described in Section 4.2. We found that *design space explorers* (*M=11.23, SD=0.91*) scored higher than the *conservative designers (M=10.62, SD=0.98)* (see Figure 2(b)*)*. The Mann-Whitney U test showed a significant difference (*p=0.046, U=323.5, Z=1.968, r=0.297*) among the clusters. Further, analysis revealed that 10 of 20 *conservative designers* (50%) and 19 of 24 *design space explorers* (79.2%) submitted an *AllS-Low Runoff* designs; the remainder submitted sub-optimal designs with various constraint combinations. 10 *conservative designers* (50%) and three *design space explorers* (12.5%) submitted designs that achieved low runoff and met the *cost criterion*; 1 *design space explorers* pair (4.2%) submitted designs that only met *blocks criteria*; and one *design space explorers* pair (4.2%) submitted a design that met both the *cost* and *blocks criteria* but had *High Runoff*. Overall, only four *conservative designer pairs* (20%), while 11(45.8%) *design space explorer pairs* submitted the optimal design achieving 12 points.

6. Discussion

Our results demonstrate significant learning gains (p < 0.001, see Section 5.1), indicating that the WRC curriculum effectively facilitated students' understanding of science, computational thinking, and engineering design concepts. However, only 29 pairs (65.9%) submitted a design meeting all constraints with low runoff, suggesting that while most students grasp the primary concepts of the WRC curriculum, several struggle to apply these concepts effectively in real-life problem-solving scenarios.

Markov analysis revealed students' diverse approaches in their search for optimal design solutions. Students struggled to integrate multiple constraints into their decision-making process– for instance, the loop between *CostS-Low Runoff* and *AllS-Low Runoff* suggests that students' attempts to optimize one constraint often came at the expense of another– a trend also seen in prior research (Zhang et al, 2019; Purzer et al, 2015). While trade-offs are inherent in design tasks, the WRC curriculum does not explicitly emphasize the role of trade-off analysis in the problem-solving process. Incorporating this analytical perspective into the curriculum could help students develop critical design thinking skills.

Analysis of the submitted designs reveals that *design space explorers* indeed reported better design solutions than *conservative designers*. More *design space explorers* submitted an *AllS-Low runoff* solution, sometimes also the optimal solution, as reported in section 5.4. This finding also supports the claim that to optimize a design truly, engineers must sufficiently explore the possible design space (Montgomery et al, 2020) and underscores the need to encourage such exploration in design activities. We can assert that the patterns observed are likely intentional design changes rather than random exploration as if students' actions were random, optimal solutions would be unlikely or take considerably longer to reach as there are 6^{12} possible solutions to this engineering design challenge.

Additionally, we found that students' final designs prioritized low runoff and the *cost criterion,* values that are visually emphasized on the design interface*,* while overlooking the *blocks criteria*. "Informed" designers use words and graphics when considering plans or justifying decisions (Purzer et al, 2015). Our findings support this, and we conjecture that displaying additional criteria (e.g., the *blocks criteria*) within the ED interface may encourage students to focus more holistically on all design constraints. In future studies, we will incorporate this to display the design criteria on the interface.

Some students generated better solutions than the ones they reported as their final submissions. Our previous investigations (Zhang et al, 2019) found that fewer than 10% of students reported their best design. Recognizing this, we added a “design history” table in the simulation environment, allowing students to compare their current designs to ones they generated earlier. We conjecture that this feature supported students in identifying their best design, as 15 pairs (34.1%) submitted the “optimal design”, which was the best design they generated. This finding suggests that supporting the recording and reviewing of design solutions can scaffold students’ optimization processes.

While the results are specific to the WRC curriculum, our approach to analyzing design strategies is generalizable. The specific states depend on the problem statement and associated design constraints, the methods for categorizing solutions, and identifying strategies from sequences of design solutions are broadly applicable. For other multi-constraint challenges (e.g., Du et al., 2025; Montgomery et al., 2020; Bowen et al., 2016), solutions can similarly be divided into states and analyzed for strategy patterns. Future work will test and validate these methods across other engineering design and open-ended learning environments.

7. Conclusion and Future Work

This paper presents an effective approach to unpacking the problem-solving behaviors of middle school students by analyzing the sequences of their generated designs. This contributes to the limited research on the design behaviors of middle school students. Furthermore, the WRC curriculum is unique as it combines Science, CT, and ED while studying the relationships between science and engineering design in a manner proposed in the NGSS curricula (Lilly et al, 2022). The results demonstrate the effectiveness of the WRC curriculum and emphasize the need for scaffolding exploration to successfully engage with Engineering Design activities.

This research has implications for incorporating Engineering Design tasks into STEM+C learning environments. Our findings suggest that students require support in performing tradeoff analysis and learning to compare and evaluate their design solutions. One potential way to assist with tradeoff analysis in these exploratory environments is to display students' design sequences and the design constraints that must be met on the interface. Students tend to focus more on the constraints visible on their screens. Future work should concentrate on effectively scaffolding students to systematically explore design solutions.

This work has limitations. The design histories used in this analysis did not capture intermediate actions between tests, so the Markov analysis misses nuances of students’ problem-solving strategies, such as reasons for specific changes from one design solution to the next. Although clustering analysis suggested distinct groups and FOMM revealed consistent patterns, further validation of problem-solving strategies will require analysis of student conversations during the design sessions. As this was a retrospective analysis, further research is needed to assess how our approach can improve instruction and scaffolding in classroom settings. In the future, we aim to leverage online analysis of students’ design solutions to identify struggling students and to provide targeted support for those using less effective exploration strategies during the design process.

Acknowledgements

We thank the National Science Foundation (#1742195), participating students, and teachers.

References

Ali, M., & Tse, A. W. C. (2023). Research Trends and Issues of Engineering Design Process for STEM Education in K-12: A Bibliometric Analysis. International Journal of Education in Mathematics, Science and Technology, 11(3), 695-727.

Basu, S., McElhaney, K. W., Rachmatullah, A., Hutchins, N. M., Biswas, G., & Chiu, J. (2022). Promoting computational thinking through science-engineering integration using computational modeling. In Proceedings of the 16th International Conference of the Learning Sciences-ICLS 2022, pp. 743-750. International Society of the Learning Sciences.
Bowen, B. D., DeLuca, V. W., & Franzen, M. M. S. (2016). Measuring how the degree of content knowledge determines performance outcomes in an engineering design-based simulation environment for middle school students. Computers & Education, 92, 117-124.
Du, H., Zhu, G., Xing, W., & Xie, C. (2025). What Distinguishes Students' Engineering Design Performance: Design Behaviors, Design Iterations, and Application of Science Concepts. Journal of Science Education and Technology, 34(2), 314-326.
Gatta, R., Vallati, M., Lenkowicz, J., Rojas, E., Damiani, A., Sacchi, L., ... & Valentini, V. (2017, December). Generating and comparing knowledge graphs of medical processes using pMineR. In Proceedings of the 9th Knowledge Capture Conference (pp. 1-4).
Goldstein, M. H., Mejia, C. V., Adams, R., Purzer, Ş., & Zielinski, M. (2016, October). Developing a measure of quality for engineering design artifacts. In 2016 IEEE Frontiers in Education Conference (FIE) (pp. 1-7). IEEE.
Hutchins, N. M., Basu, S., McElhaney, K., Chiu, J., Fick, S., Zhang, N., & Biswas, G. (2021). Coherence across conceptual and computational representations of students' scientific models. Computer-supported collaborative learning.
Kelly, G. J., & Cunningham, C. M. (2019). Epistemic tools in engineering design for K-12 education. Science Education, 103(4), 1080-1111.
Lilly, S., McAlister, A. M., Fick, S. J., Chiu, J. L., & McElhaney, K. W. (2022). Elementary teachers' verbal supports of science and engineering practices in an NGSS-aligned science, engineering, and computational thinking unit. Journal of Research in Science Teaching, 59(6), 1035-1064.
McElhaney, K. W., Zhang, N., Basu, S., McBride, E., Biswas, G., & Chiu, J. L. (2020, January). Using computational modeling to integrate science and engineering curricular activities. In M. Gresalfi & IS Horn (Eds.). The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences (ICLS) 2020 (Vol. 3).
McHugh, M. L. (2012). Interrater reliability: the kappa statistic. Biochemia medica, 22(3), 276-282.
Montgomery, R., Greenwald, E., Crane, S., Krakowski, A., Barber, J.: Operationalizing optimization in a middle school virtual engineering internship. Journal of Science Education and Technology 29(3), 409–420 (2020). https://doi.org/https://doi.org/10.1007/s10956-020-09826-8
National Research Council, Division of Behavioral, Board on Science Education, & Committee on a Conceptual Framework for New K-12 Science Education Standards. (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. National Academies Press.
NGSS Lead States. (2013). Next generation science standards: For states, by states. National Academies Press.
Purzer, Ş., Goldstein, M. H., Adams, R. S., Xie, C., & Nourian, S. (2015). An exploratory study of informed engineering design behaviors associated with scientific explanations. International Journal of STEM Education, 2, 1-12.
Radloff, J., & Capobianco, B. M. (2021). Investigating elementary teachers' tensions and mitigating strategies related to integrating engineering design-based science instruction. Research in Science Education, 51, 213-232.
Srivastava, N., Fan, Y., Rakovic, M., Singh, S., Jovanovic, J., Van Der Graaf, J., ... & Gasevic, D. (2022, March). Effects of internal and external conditions on strategies of self-regulated learning: A learning analytics study. In LAK22: 12th international learning analytics and knowledge conference (pp. 392-403).
Wendell, K. B., Wright, C. G., & Paugh, P. (2017). Reflective decision-making in elementary students' engineering design. Journal of Engineering Education, 106(3), 356-397.
Xing, W., Pei, B., Li, S., Chen, G., & Xie, C. (2023). Using learning analytics to support students' engineering design: the angle of prediction. Interactive Learning Environments, 31(5), 2594-2611.
Yu, K. C., Wu, P. H., & Fan, S. C. (2020). Structural relationships among high school students' scientific knowledge, critical thinking, engineering design process, and design product. International Journal of Science and Mathematics Education, 18(6), 1001-1022.
Zhang, N., Biswas, G., Chiu, J. L., & McElhaney, K. W. (2019). Analyzing students' design solutions in an NGSS-aligned earth sciences curriculum. In Artificial Intelligence in Education: 20th International Conference, AIED 2019, Chicago, IL, USA, June 25-29, 2019, Proceedings, Part I 20 (pp. 532-543). Springer International Publishing.
Zhou, N., Pereira, N., Chandrasegaran, S., George, T. T., Booth, J., & Ramani, K. (2021). Examining middle school students' engineering design processes in a design workshop. Research in Science Education, 51, 617-646.