Analyzing Students' Design Solutions in an NGSS-Aligned Earth Sciences Curriculum

Ningyu Zhang[1(✉)], Gautam Biswas[1], Jennifer L. Chiu[2], and Kevin W. McElhaney[3]

[1] Department of EECS, Institute for Software Integrated Systems, Vanderbilt University, 1025 16th Avenue South, Nashville, TN 37212, USA
{ningyu.zhang,gautam.biswas}@vanderbilt.edu
[2] Department of Curriculum, Instruction and Special Education, University of Virginia, PO Box 400273, Charlottesville, VA 22904, USA
jlc4dz@virginia.edu
[3] Education Division, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025, USA
kevin.mcelhaney@sri.com

Abstract. This paper analyzes students' design solutions for an NGSS-aligned earth sciences curriculum, the Playground Design Challenge (PDC), for upper-elementary school (grade 5 and 6) students. We present the underlying computational model and the user interface for generating design solutions for a school playground that has to meet cost, water runoff, and accessibility constraints. We use data from the pretest and posttest assessments and activity logs collected from a pilot study run in an elementary school to evaluate the effectiveness of the curriculum and investigate the relations between students' behaviors and their learning performances. The results show that (1) the students' scores significantly increased from pretest to posttest on engineering design assessments, and (2) students' solution-generation and testing behaviors were indicative of the quality of their design solutions as well as their pre-post learning gains. In the future, tracking such behaviors online will allow us to provide adaptive scaffolds that help students improve on their engineering design solutions.

Keywords: Technology-enhanced learning · NGSS · Engineering design · Learning analytics

1 Introduction

Design activities provide learners a supportive, authentic, and effective context to experiment with and develop an understanding of real-world scenarios using models of scientific processes [12,15]. Design-based learning activities, especially complex design problems, have shown great potential and promise in benefiting K-12 students' learning [5,6,15]. The Next Generation Science Standards

S. Isotani et al. (Eds.): AIED 2019, LNAI 11625, pp. 532–543, 2019.
https://doi.org/10.1007/978-3-030-23204-7_44

(NGSS) of the United States include engineering disciplinary core ideas and practices within the three-dimensional performance expectations (PEs) of disciplinary core ideas, science and engineering practices, and crosscutting concepts as early as kindergarten [7]. There are also efforts from the engineering and science education community to promote engineering design activities in elementary classrooms to strengthen science education [15]. However, previous studies have also highlighted the challenges and barriers to integrating engineering concepts and practices into elementary school curricular settings [5,9].

This paper reports on the Playground Design Challenge (PDC), an NGSS-aligned curricular unit for upper elementary school students (age 11 to 12). PDC integrates the Earth science and engineering domains through (1) scientific investigations that involve physical experiments on the absorption of different surface materials, (2) building conceptual models to understand the concepts of water absorption and runoff after a rainfall; (3) an engineering design challenge, where students can design playground models that meet specified constraints, and evaluate the construction cost and total water runoff of a designed playground [3]; and (4) use of computational models implemented in NetsBlox [1] that enable students to test different design solutions.

In the rest of this paper, we present the learning environment and the underlying computational model used in the PDC. We describe the data collected from a pilot study to evaluate the effectiveness of the curriculum and investigate the relations between students' behaviors and their pre-post learning gains. More specifically, we investigated the following research questions:

RQ 1: How effective was the intervention in improving students engineering design proficiency?

RQ 2: How well did the students' engineering design solutions correlate with their pre- to posttest learning gains?

RQ 3: How did students' behaviors of exploring the *problem space* and generating their engineering designs align with the performances on the NGSS PEs?

2 Engineering Design and the K-12 Curricula

Engineering design involves complex cognitive processes such as (1) understanding the problem, (2) generating ideas, (3) learning new concepts necessary for solving problems, (4) developing and testing models, and (5) analyzing and revising solutions [9]. Design-based instruction is more accessible to elementary school students as younger learners tend to have "less apprehension toward design challenges" compared to elder learners [15, p. 515]. In addition, design activities have great potential and promise to benefit science learning because scientific scenarios can be contextualized into compelling design problems [5,6,15].

Whereas science learning through problem-solving has received a lot of attention in secondary school curricula *e.g.,* [6] there is much less focus on *science-through-design* learning for younger pupils [15] (exceptions are Penner *et al.* [12] and Wendell *et al.*'s [15] work). Penner *et al.*'s study with third-grade students involved designing models of the human elbow. Students engaged in a series of design-related activities such as building, testing, and evaluating models. The students then used the elbow models to explore the biomechanics of the human body [12]. Wendell *et al.* implemented a LEGO™ design challenge for elementary-grade students that created a synergy between science learning knowledge and engineering design [15]. Both studies demonstrated how students' problem-solving processes in a *problem space* [10] could provide engaging teaching and learning strategies.

3 Methods

3.1 The Engineering Design Environment

We created a computational model to simulate the effect of rainfall on water runoff from a playground to surrounding areas. The total runoff rate is calculated by the Rational Equation, a widely-applied method in hydraulic engineering to estimate the peak discharge of a small watershed [14]. The equation for peak discharge volume is $Q = c \times i \times A$, where c is a unit-less runoff ratio, i is the rain intensity, and A is the drainage area. To make the equation more understandable to elementary school students, we simplified the equation by assuming that the playground had a unit area, thereby eliminating the area variable from the model. As a result, the runoff coefficient is interpreted as the amount of discharge per unit of rainfall intensity (measured in inches) on the playground.

In this pilot study, students were provided access to a pre-built, interactive computational model implemented in NetsBlox [1], which they used to construct and test playground designs[1]. Students could combine seven surface materials for constructing their playground: (1) concrete, (2) natural grass, (3) artificial turf, (4) engineered wood chips, (5) sand, (6) rubber tiles, and (7) poured rubber. As part of their design task, students chose materials that were appropriate for the different parts of the playground (this was specified as requirements for specific play areas, e.g., soccer field, basketball court, swing sets, etc.) and met the runoff and cost constraints; in addition, they had to make sure that the field was wheelchair-accessible. The cost, runoff ratio, and accessibility of each material were provided to the students to help them design the playground.

Students constructed their playground by clicking on the squares and selecting from the seven available materials. When students chose a material, the square's look reflected the choice of material, and its cost was added to the total playground cost. The total runoff from the playground for a specified amount of rainfall was also updated using the Rational Equation.

The left part of Fig. 1 shows a playground design built with materials such as natural grass, artificial turf, sand, concrete, and rubber tiles to allow for a soccer field (four squares), a basketball court (two squares), and a play area with swings and other equipment (two squares). The three icons on the top of the UI on Fig. 1 represent the control buttons for the simulation. Students could click on

[1] Students did not program the computational model in this pilot study, however, we have added programming activities in NetsBlox for future studies.

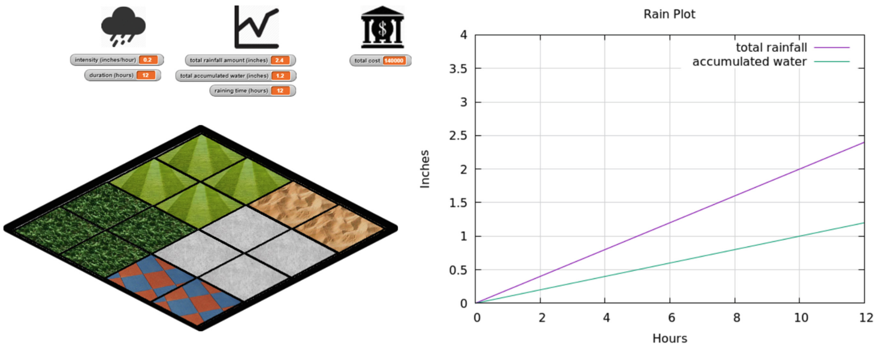

Fig. 1. The user interface of the playground design and a rain plot.

the dark cloud icon to open a dialog box to select the intensity and duration of the rain. After a simulation run, students could view a plot of the results (rainfall amount and runoff by the hour), and check the cost of the current playground under the bank icon.

The system logged five types of actions as students interacted with the computational model: (1) adding/removing the surface material for a square; (2) resetting all squares on the playground to the initial empty state, and (3–5) clicking on the 3 control buttons. The values of the model variables (*e.g.*, the choice of the surface materials, the total cost, and runoff rate) were also logged for *post hoc* analyses.

3.2 Playground Design Criteria and Scoring

The students were informed that a satisfying playground design must meet three criteria: (1) runoff $\leq$ 0.5 in. after 1.2 in. of total rainfall in 4 h, (2) cost $\leq$ \$200,000 for the playground, and (3) having sufficient accessibility for students in wheelchairs. The accessibility criterion was not quantified in the design specifications, but for our *post hoc* analysis of the students' designs, we assigned scores of 1.0, 2.0, and 3.0 to low, medium, and high accessibility materials. Because the values associated with the three design criteria had widely different scales, we applied a simple transformation to each criterion to reduce them to a value between 1.0 and 5.0 to ensure that each criterion was given equal weight in our assigned evaluation score for a student's design. Specifically, because the playground cost varied between \$40,000 and \$600,000, we applied an inverse linear scaling to convert the actual playground cost into a score in the range [1.0, 5.0], where 1.0 represented \$600,000 and 5.0 represented \$40,000. Similarly, the runoff values were scaled to the same range with 1.0 representing a 0.96-in. runoff design (the maximum possible) and 5.0 representing a 0.24-in. runoff (the minimum possible) after the 1.2-in. rainfall.

We then used the mean of the 3 sub-scores as the score of a playground design. Figure 2 presents a visualization of a baseline design that just meets all of the criteria, *i.e.*, a design costing $200,000, resulting in 0.5 in. of runoff after a 1.2-in. rainfall, and having a medium level of accessibility. The score computed for this baseline design is 3.4 (the mean of the runoff score of 3.35, cost score of 3.85, and accessibility score of 3). The students were not aware of this scoring system while they worked on their designs. Instead, they directly compared the cost of the playground and the amount of the runoff for a number of designs and then selected what they argued was their best solution.

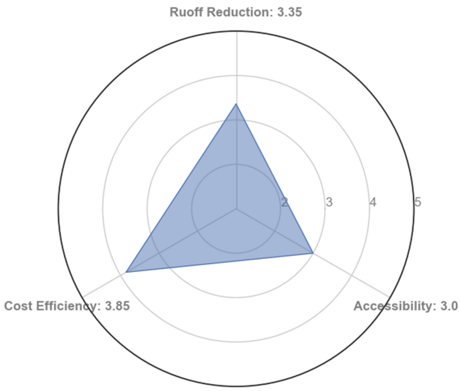

Fig. 2. The scores of the baseline design calculated by post hoc analysis

3.3 Assessment of Integrated Science and Engineering Proficiency

We designed a summative assessment that included 3 tasks measuring students' proficiency with NGSS upper elementary engineering design Performance Expectations [7]. One task assessed students' ability to define a design problem (3-5-ETS1-1). The second and third tasks assessed students' ability to generate and compare multiple possible solutions (3-5-ETS1-2) [8]. The assessment modality included multiple choice and constructed response questions. All three tasks were designed around the scientific concepts of water runoff. Task rubrics rewarded the extent to which students could make valid engineering decisions and whether these decisions were informed by the underlying scientific concept of water absorption and runoff. Sample questions and more detailed discussions of the development of the 3-dimensional assessment, its alignment to the NGSS, and the description of the grading rubrics have been presented in [3,8].

A total of 397 students (123 fifth-graders and 274 sixth-graders) from an upper-elementary school in the United States participated in the 4-week pilot study (about 1 h per school day). The study was led by science teachers with researchers playing the role of observers. The school district's STEM coordinator and one participating teacher were closely involved in the development and implementation of the curriculum. However, the teachers had not taught the PDC curriculum prior to this implementation. The summative assessment was administered as a pre- and posttest at the beginning and the end of the study.

4 Results and Discussions

4.1 Learning Gains from the Curriculum Unit

The pre- and posttest scores of a subset of 107 students were graded at the time of this analysis. We did not include students who missed a pretest or a posttest, leaving us with 88 students. We confirmed with a Kolmogorov-Smirnov test [4] that the pre-post scores were not normally distributed, and then used the non-parametric Wilcoxon rank-sum test [4] to examine if the differences in pre- to posttest scores were significant. Table 1 reports the test statistics of the overall scores and their breakdown.

Table 1. Learning gains (N = 88).

Test	Points	Pre score (std)	Post score (std)	p-value	z-score	Effect size
Total	18	4.72 (3.52)	6.50 (3.61)	<0.001	4.68	0.35
Def. problem	6	1.44 (1.11)	2.03 (1.33)	<0.001	3.77	0.28
Gen. solution	4	1.59 (1.45)	2.02 (1.52)	0.012	2.50	0.19
Comp. solutions	8	1.68 (1.78)	2.44 (1.76)	0.008	2.67	0.20

The students' learning performance showed statistically significant improvements in all aspects. This helps answer RQ 1, *i.e.*, that there appeared to be a positive association between the curriculum and the students' improved proficiency in engineering design tasks. However, the effect sizes were small, and there was a considerable gap between the posttest scores students attained the maximum possible score. Therefore, there is room for students to improve their engineering design abilities, which we hope to achieve by refining the current curriculum. This result also matches the literature that engineering design is challenging for elementary school students [15].

4.2 Playground Design Behaviors and Design Scores

As discussed in Sect. 3.1, the system logs five types of actions as students experiment and design their playgrounds. During the study, it recorded a total of 79,003 actions from 357 students. In this paper, we focus on two types of measures of the log data that relate to evaluating design solutions: (1) the number of tests conducted by a student; and (2) the scores assigned to a student-generated playground design. Both measures indicate how the students searched the solution space [10] and how well the generated solutions met the design criteria.

Table 2 presents the definition and descriptive statistics related to the number of tests the students conducted and the solution they chose. The relatively large variance in the number of test actions can be explained by classroom observations that some students worked in pairs to generate these solutions. This was expected because students were encouraged to work with each other and discuss their solutions with others in the classroom.

Table 2. Descriptive statics of test data.

Variable name	Description	Mean (std)	Range
Num test	Number of tested designs	7.64 (7.19)	1–36
Satisfy designs	Number of tested designs satisfying all criteria	3.28 (4.53)	0–27
Best score	Highest score of all tested designs	3.86 (0.15)	2.85–4.03
Last score	Score of the last (temporally) tested design	3.76 (0.17)	2.85–4.03
Submitted score	Score of the design submitted to WISE	3.77 (0.14)	3.42–3.97
Score diff	Difference between submitted and best scores	0.08 (0.18)	−0.30–0.47

Students submitted their final playground design to the Web-based Inquiry Science Environment (WISE) [13], where they also participated in a number of instructional activities. The submitted designs were scored by the method discussed in Sect. 3.2. A negative value for the submitted score difference (the last row of Table 2) implies that some students submitted a design that was better than what they tested during their computational modeling experiments. This discrepancy can be partly explained by the fact that students collaborated for some of the time and the solution reported may have resulted not from individual work but the collaboration, which produced better solutions than the individual efforts. On the other hand, classroom observations and interviews also indicated that some students arbitrarily reported designs that they thought looked good, although they did not actually test these solutions. The second situation echoed reports in the literature that students' focus during design activities may be diverted by personal aesthetics [5].

We compared the scores of the students' *submitted* designs to the highest scores of *tested* designs and found that less than 10% of the students reported their best design on WISE. A Mann-Whitney U-test [4] showed a significant difference in the scores. The average submitted score was 3.77 (stdev $= 0.14$) and the average best solution score generated in the NetsBlox environment was 3.86 (stdev $= 0.15$). This difference was significant (p-value < 0.001) with a large effect size of 3.48. This result indicates that although the students were able to generate satisfying design solutions, they did not make much of an attempt to compare the different solutions they had generated. More importantly, the difference in the reported solution and the best solution provides insight into students' understanding and learning to generate optimal playground design solutions. We discuss the implications in Sect. 4.3.

4.3 Correlation Analyses

Table 3 reports the correlation coefficients (Spearman's ρ) of the performance and behavioral measures from the 88 students whose pretest and posttest scores were available. Statistically significant correlations are marked with *s. We present a few observations from the correlation analysis and discuss how they can help answer research questions 2 and 3.

Table 3. Correlation coefficients of measures (*: p-value < 0.05, **: p-value < 0.01).

	Pre score	Post score	Learning gain	Num test	Satisfy designs	Best score	Last score	Sbmtd. score
Post score	**0.64****							
Learning gain	−0.25*	**0.53****						
Num test	0.12	0.13	0.05					
Satisfy designs	0.12	0.00	−0.13	**0.75****				
Best score	0.17	0.06	−0.11	**0.80****	**0.81****			
Last score	0.06	0.04	−0.10	**0.66****	**0.65****	**0.53****		
Submitted score	−0.09	−0.09	**0.49****	**0.31***	0.17	−0.04	0.23*	
Score diff	0.11	0.04	**−0.41****	0.23*	0.06	0.06	0.05	0.18

First, the students' pre- and posttest scores are highly correlated ($\rho = 0.64$). This is an expected result—studies [2,11] suggest that a learner's prior knowledge in a domain facilitates further learning in the domain. Surprisingly, the weak negative correlation between pretest score and learning gains ($\rho = -0.25$) implies those who had high prior knowledge did not learn as much from the intervention. The small negative correlation between learning gains and the playground design scores ($\rho = -0.11$) implies that the students learned about design criteria, but may have not applied them in an effective way to generate their design solutions. However, the correlation is not significant, implying there may be no true effect between the two variables.

Second, the learning gain is correlated with the submitted design scores ($\rho = 0.49$). We expected such a correlation because we believed that the students' performance in the engineering design activities should contribute to their improvement of the engineering proficiency (as evaluated by the pre-post assessments). This observation provides insights into RQ 2 that the students' engineering design solutions are indicative of their learning gains.

Third, the number of tests conducted by the students is correlated with (1) the number of satisfying designs ($\rho = 0.75$), (2) the best design scores ($\rho = 0.80$), and (3) the submitted design scores ($\rho = 0.31$). Additionally, the number of satisfying designs also correlated with the highest design scores ($\rho = 0.81$). It suggests that the students who committed more effort on systematically creating and testing design solutions were more likely to find better playground designs.

Fourth, the *difference* between the submitted design scores and the best design scores is moderately and negatively correlated with the learning gain ($\rho = -0.41$). In addition, when we analyzed the correlation of this variable and the learning gains of each individual assessment items (*i.e.*, defining problem, generating solutions, and comparing solutions – results not reported in Table 3 due to the space limitations), we found that the learning gains for the *comparing solutions* sub-task is strongly and negatively correlated with the score difference ($\rho = -0.72$). This evidence suggests that students' ability to discern better design solutions in the learning environment is strongly indicative of the NGSS PE of comparing solutions (3-5-ETS1-2).

Fifth, the submitted scores had a larger correlation to the score of the *last* tested design ($\rho = 0.23$) than the *best* design ($\rho = -0.04$). Because the submitted scores seem to be independent of the best design scores, it is reasonable to believe that a large number of students simply reported the results of their last generated design solution rather than the best design solutions. Nonetheless, these observations provide some evidence to answer RQ 3 that students' learning behaviors and performances directly link to the NGSS engineering performance.

4.4 A Case Description

In this subsection, we present a case study using the log data from one student to illustrate his/her playground design processes. The student was among the most successful students in the pilot study based on their learning improvement and design performances. The student tested the designs 29 times (at the 98^{th} percentile, abbreviated as % later), and 13 tested designs satisfied all design criteria (95%). The student had the highest design score of 3.92 (67%), which is also the design submitted to WISE (85%). The student's overall pretest score, posttest score, and standardized learning gain were 5 points, 8 points, and 0.42 (59%, 99%, and 90%, respectively). Figure 3 provides a visualization of the students' playground design projected onto a 3-dimension space. The three axes of the figure correspond to the runoff, cost, and accessibility aspects of the design criteria. Each dot on the 3-D plot marks a tested design. The shaded region stands for a satisfying solution space, *i.e.*, all dots contained in the solution space mark a satisfying design.

The student's initial design used poured rubber (the most expensive material) on 4 squares and had a total cost of \$255,000, failing to satisfy the cost criterion. On the second try, the student replaced poured rubber on 2 squares with less expensive materials and made a satisfying design. Despite succeeding early on, the student continued to explore additional solutions in an apparent effort to further improve the solution. The student tried other designs using more concrete, a less absorbent material that caused more runoff, resulting in a few designs that again failed the runoff criteria (designs 3–7). After addressing the runoff problem, the student made the best design at the 10^{th} attempt and kept experimenting. Later, the student replaced half of the concrete squares with natural grass, which in turn caused the playground not being accessible anymore (designs 14–16). Then the student tried a new design with artificial

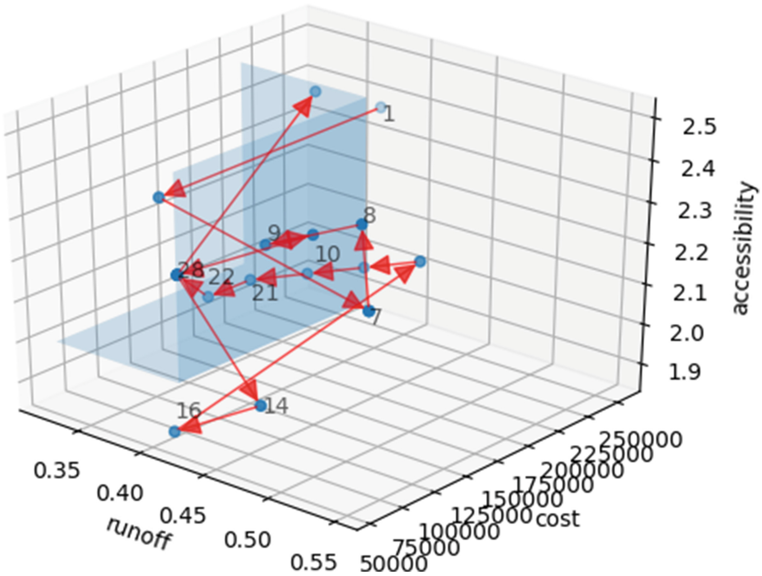

Fig. 3. A student's playground design projected on a 3-dimension space.

turf and concrete, raising the total cost over the limit again (No. 17). Finally, after exploring a few other inexpensive designs consisting most of natural grass and concrete, the student created a satisfying final solution. By this time, the student had experimented with all 7 surface materials.

This case study presents the trajectory of a successful designer who also achieved high learning gains. More importantly, it shows how we can derive features such as (1) the transitions between non-satisfying designs and satisfying designs and (2) the changes between the designs (visualized as the arrows in Fig. 3). These features will provide a great opportunity to use data-driven methods to characterize students' learning behaviors and provide feedback to help them improve their design proficiency over time.

5 Conclusions

In this paper, we introduced the computational model and the learning activities that the students engaged in the Playground Design Challenge. We presented results from the data collected from a pilot study and discussed how the students' behaviors in the design activity could influence the performance of the design, which in turn linked to their learning performance as evaluated by NGSS-aligned pre-post assessments. For future work, (1) we have integrated the computational modeling activities in the latest version of the curriculum and planned to investigate the synergistic effect between the scientific modeling activities and engineering design activities. (2) To assist the engineering design process, we plan to add a function to record each tested design that can ease comparing these solutions without memorizing all of them. (3) We are also building tools to analyze students' log data online and provide adaptive scaffolding with methods outlined in our previous work, *e.g.*, [16,17].

Acknowledgments. This material is based upon work supported by the National Science Foundation under Grant No. DRL-1742195. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank Ron Fried, Reina Fujii, Satabdi Basu, and James Hong for their work.

References

1. Broll, B., et al.: A visual programming environment for introducing distributed computing to secondary education. J. Parallel Distrib. Comput. **118**, 189–200 (2018)
2. Chambers, S.K., Andre, T.: Gender, prior knowledge, interest, and experience in electricity and conceptual change text manipulations in learning about direct current. J. Res. Sci. Teach. Off. J. Natl. Assoc. Res. Sci. Teach. **34**(2), 107–123 (1997)
3. Chiu, J., McElhaney, K.W., Zhang, N., Biswas, G., Fried, R., Basu, S., Alozie, N.: A principled approach to NGSS-aligned curriculum development integrating science, engineering, and computation: a pilot study. Paper presented at the 2019 NARST Annual International Conference (2019)
4. Conover, W.J., Conover, W.J.: Practical Nonparametric Statistics. Wiley, New York (1980)
5. Kafai, Y.B., Ching, C.C.: Affordances of collaborative software design planning for elementary students' science talk. J. Learn. Sci. **10**(3), 323–363 (2001)
6. Kolodner, J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., Puntembakar, S.: Putting a student-centered learning by design curriculum into practice: lessons learned. J. Learn. Sci. **12**(4), 495–547 (2003)
7. NGSS Lead States: Next Generation Science Standards: For States, By States (2013)
8. McElhaney, K.W., Basu, S., Wetzel, T., Boyce, J.: Three-dimensional assessment of NGSS upper elementary engineering design performance expectations. Paper presented at the 2019 NARST Annual International Conference (2019)
9. Mehalik, M.M., Doppelt, Y., Schuun, C.D.: Middle-school science through design-based learning versus scripted inquiry: better overall science concept learning and equity gap reduction. J. Eng. Educ. **97**(1), 71–85 (2008)
10. Newell, A., Simon, H.A.: Human Problem Solving. Prentice-Hall Englewood Cliffs, Upper Saddle River (1972)
11. Osman, M.E., Hannafin, M.J.: Effects of advance questioning and prior knowledge on science learning. J. Educ. Res. **88**(1), 5–13 (1994)
12. Penner, D.E., Lehrer, R., Schauble, L.: From physical models to biomechanics: a design-based modeling approach. J. Learn. Sci. **7**(3–4), 429–449 (1998)
13. Slotta, J.D., Linn, M.C.: WISE Science: Web-Based Inquiry in the Classroom. Teachers College Press, New York (2009)
14. Thompson, D.B.: The rational method. Civil Engineering Department Texas Tech University, pp. 1–7 (2006)

15. Wendell, K.B., Rogers, C.: Engineering design-based science, science content performance, and science attitudes in elementary school. J. Eng. Educ. **102**(4), 513–540 (2013). https://doi.org/10.1002/jee.20026
16. Zhang, N., Biswas, G.: Understanding students' problem-solving strategies in a synergistic learning-by-modeling environment. In: Penstein Rosé, C., et al. (eds.) AIED 2018. LNCS (LNAI), vol. 10948, pp. 405–410. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93846-2_76
17. Zhang, N., Biswas, G., Dong, Y.: Characterizing students' learning behaviors using unsupervised learning methods. In: André, E., Baker, R., Hu, X., Rodrigo, M.M.T., du Boulay, B. (eds.) AIED 2017. LNCS (LNAI), vol. 10331, pp. 430–441. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-61425-0_36