Using Artificial Intelligence Teaching Assistants to Guide Students in Solar Energy Engineering Design in Response to COVID

Rundong Jiang, Xiaotong Ding, Dylan Bulseco, and Charles Xie

Institute for Future Intelligence, Natick, MA 01760

Curriculum & Instruction

Keywords: engineering design; artificial intelligence; renewable energy; feedback; COVID-19

Author Note

Corresponding author: Rundong Jiang (15 Village Rock Ln, Apt 8, Natick, MA 01760; rundong@intofuture.org)

Abstract

Engineering projects, such as the design of a solar farm, engage students in addressing real-world challenges by learning and applying geoscience knowledge. To improve their designs, students need frequent and informative feedback as they iterate. However, teacher attention may be limited or inadequate, both during COVID and beyond. We present Aladdin, a web-based CAD/CAE platform for engineering design with a built-in artificial intelligence teaching assistant (AITA). We also present two curriculum units (Solar Energy Science and Solar Farm Design), where students explore the Sun-Earth relationship and optimize the energy output and yearly profit of a solar farm with the help of the AITA. We tested the software and curriculum units with over 100 students in two Midwestern high schools. Pre- and post-test data showed improvements in understanding of science concepts and self-efficacy in engineering design, with noticeable differences between the two schools. We identified common patterns in student interactions with the AITA that led to either improvements or misconceptions. Interviews revealed students' values and preferences when receiving feedback. Our findings suggest that AITAs may be helpful as an additional feedback mechanism for geoscience and engineering education. Future research should focus on improving the usability of the software and providing multiple types of feedback to promote inclusive and equitable use of AI in education.

Using Artificial Intelligence Teaching Assistants to Guide Students in Solar Energy Design During COVID

INTRODUCTION

There is an increasing demand to integrate engineering design with geoscience education. At the K-12 level, the Next Generation Science Standards (NGSS) listed 7 earth and space science (ESS) performance expectations that incorporate engineering practices. For example, high school students are expected to “evaluate competing design solutions for developing, managing, and utilizing energy and mineral resources based on cost-benefit ratios” (NGSS Lead States, 2013). In addition, the K-12 science faculty, including ESS educators, share the responsibility to address 14 separate NGSS performance expectations for engineering design, such as “design[ing] a solution to a complex real-world problem by breaking it down into smaller, more manageable problems that can be solved through engineering” (NGSS Lead States, 2013). According to the National Research Council's *A Framework for K-12 Science Education*, on which the NGSS are based, a major advantage of integrated science and engineering education is that “[f]rom a teaching and learning point of view, it is the iterative cycle of design that offers the greatest potential for applying science knowledge in the classroom and engaging in engineering practices” (NRC, 2012). The interactivity of and repeated involvement in engineering design projects may also help trigger and maintain students’ situational interest in ESS (van der Hoeven Kraft, 2017).

Of all engineering design projects within an ESS context, renewable energy engineering may be one of the most familiar to a K-12 audience. Take solar energy engineering - the design and deployment of solar power systems - for an example. Prior research suggests that while an overwhelming majority of students reported some familiarity with the concept of solar panels and many reported seeing them in their everyday lives, much fewer could use ESS knowledge such as solar angles to explain what time of day solar panels worked best (Kishore & Kisiel,

2013). Therefore, a solar energy design project can both relate to students' personal experiences with solar energy and reinforce their ESS knowledge through repeated application in an iterative design process. For example, how can the design of utility-scale solar panel arrays - or solar farms - integrate ESS with engineering? For starters, the energy output of solar panels depends on solar irradiance, which fluctuates according to the Sun's position in the sky. An optimal solar farm design will use an appropriate tilt angle to maximize the solar insolation (the total incident solar radiation across a certain time) and thus the energy output. In addition, the exact energy output of a solar farm is dependent on a number of other factors, such as the latitude (which determines the daytime length and the Sun's relative position), the local weather (which determines the number of sunshine hours), etc., which must be accounted for in an accurate yield analysis.

The importance of solar energy engineering was further highlighted during the COVID-19 crisis. Its impact on geoscience education reached far beyond an extended absence of field-based practical opportunities and a hurried shift towards virtual learning tools and online teaching (Riggs, 2020). The pandemic also brought other critical issues into the foreground, such as the climate crisis and how both tragedies disproportionately affect marginalized communities (Bethune, 2020). Solar energy engineering can be a powerful response to this issue, with the Biden Administration signing an executive order to decarbonize the energy sector (The United States Government, 2021) and the amount of renewable energy generated reaching a record high of 28% in April 2022 (U.S. Energy Information Administration, 2022). Meeting the decarbonization goal requires a solar workforce of as many as 500,000-1,500,000 people by 2035 (U.S. Department of Energy, 2021), which serves as a reminder that engineering design projects should be integrated into regular ESS education, so that students can be prepared to apply their geoscience knowledge to mitigating global challenges such as the climate crisis.

However, teaching engineering design is hardly an easy task even in normal times, not to mention during a global pandemic. The engineering design process is iterative and improvement is incremental, meaning that students would typically require frequent feedback on their design process and product, often from either their teachers or their peers, so that they can evaluate the pros and cons of their current design, assess their application of scientific principles, and explore potential next steps. The inconvenient truth is that even in normal times, teachers are often unable to look over the shoulder of each student to provide individual feedback on each design iteration due to a lack of either time or expertise. Peer feedback may be more available but not necessarily as effective without proper training. The situation was exacerbated by the total interruption of all face-to-face interactions at the height of the pandemic, meaning that students could be left with no feedback at all during their learning.

In addition to introducing new challenges, the COVID-19 crisis also highlighted existing shortcomings in science and engineering education, especially around issues of equity and inclusion. For example, the cost of physical materials can become a barrier for students from low socioeconomic statuses, limiting their access to and success in engineering design projects. Traditional engineering projects may not be accessible for students with chronic illness or disabilities, who may rely more on virtual learning than their peers (Porter et al., 2021; Thornton et al., 2022). Also, some students may not actively seek teacher or peer feedback due to their personality or neurodiversity. In each case, the lack of alternatives may discourage certain students from developing interest or expertise in science and engineering. While remote learning is unlikely to dominate once the pandemic subsides, the necessity of an ever available virtual option has become evident to students and teachers, should the external environment ever necessitates it or any individual request it. Similarly, alternative feedback mechanisms need not replace all in-person teacher and peer feedback, but they can serve as a safety net and allow students to personalize their learning based on their diverse needs.

Recent developments in artificial intelligence (AI) have propelled a wave of educational applications in assessment, tutoring, and feedback (Goldin et al., 2017; Mirchi et al., 2020; Porter and Grippa, 2020; Afzaal et al., 2021; Darvishi et al., 2022; Hooda et al., 2022). In the field of engineering design, AI has been used in computer-aided design (CAD) and computer-aided engineering (CAE) settings (Shu et al., 2019; Yoo et al., 2021), computational geoscience (Bergey, 2020), and renewable energy engineering (Vahdatikhaki et al., 2022). There has been some exploration of its capability to assess engineering design performance (Xing et al., 2021), but little has been said about its potential as a feedback mechanism in engineering education.

To advance inclusive and equitable science and engineering education and promote student agency in developing solutions to global challenges using geoscience knowledge, we introduce 1) a virtual platform for engineering design called Aladdin (Figure 1); 2) a built-in artificial intelligence teaching assistant (AITA) capable of providing individual design feedback, and; 3) a week-long solar energy science and engineering curriculum. Aladdin is an integrated computer-aided design (CAD) and computer-aided engineering (CAE) tool for renewable energy engineering (Xie et al, 2018). The design of the AITA was informed by the field of heuristics (Gigerenzer, 2008), which has a long tradition in math teaching (Higgins, 1971; Hughes, 1974; Lucas, 1974), has been observed as a scaffolding technique for teaching assistants (Radford et al., 2014), and was viewed as a suitable solution for AI agents as well (Al-Shaery et al., 2022). The week-long curriculum consists of two units: In the first unit, “Solar Energy Science”, students explore basic ESS concepts related to solar energy engineering, such as the solar angles and projection effect. In the second unit, “Solar Farm Design”, students explore the design requirements of a solar farm, follow the engineering design process to create their own solar farm designs, and use the AITA to improve their designs. We also present our evaluation of the software and curriculum using data from a recent field study and discuss limitations and future opportunities for AITAs in geoscience education.

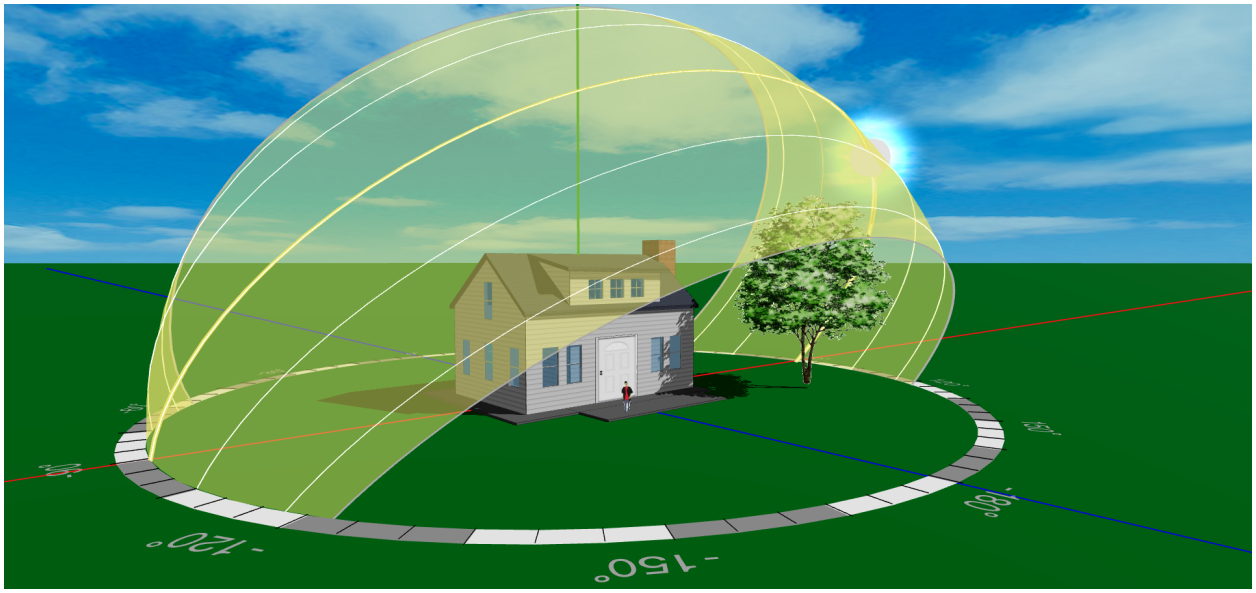


Figure 1. A screenshot of the virtual heliodon feature in the Aladdin software. The virtual heliodon visualizes the Sun's current position relative to an observer on earth and the Sun's all possible paths throughout a year for a given location. Students can input any date, time, and latitude into Aladdin, and the heliodon visualization will update automatically to reflect the change. Students can also turn on the animation feature to see the Sun move across the sky in a day.

STUDY POPULATION AND SETTING

The study took place in May 2022 in two suburban high schools in a Midwestern state. The demographics were reported in Table 1. Both schools had resumed in-person learning at the time of the study. In school 1, students sat individually and used their own laptops. In school 2, students sat in groups of 1-4 people and used school-issued Chromebooks. Two science teachers - one from each school - participated in 3 hours of professional development before implementing the solar farm design curriculum in their classrooms.

	School 1	School 2
Number of periods	4	2
Total number of students enrolled	111	31
Subject	AP Environmental Science	Physical Science
Age	15-18 (mode: 17)	16-18 (mode: 17)
Gender	Female: 45.0% Male: 39.6% Didn't report: 12.6% Prefer not to answer: 2.7%	Male: 54.8% Female: 38.7% Prefer not to answer: 6.5%
Ethnicity (only showing those with population > 5%)	White/Caucasian: 55.0% Asian/Pacific Islander: 20.7% Didn't report: 12.6%	White/Caucasian: 61.3% Multiple ethnicity: 16.1% Hispanic American : 6.5% Black or African American: 6.5%

Table 1. The demographics information of both schools that participated in the study.

MATERIALS AND IMPLEMENTATION

Before the implementation, the teachers received access to the free Aladdin software (http://intofuture.org/aladdin.html), the Solar Energy Science unit (http://intofuture.org/aladdin-solar-science.html), and the Solar Farm Design unit (http://intofuture.org/aladdin-solar-farm-design-ai.html). The units included student worksheets, teacher guides, design journals, and links to pre-made Aladdin models. Students and teachers could run Aladdin directly in the browser using their Chromebooks or laptops without having to download anything. All worksheets and surveys were also completed online using Google Suite. The teachers had editor access to all Google Docs files and could view and leave comments on student worksheets. They could also view the Google Form responses of the pre- and post-test surveys.

The full curriculum took 5-7 days to implement. On the first day, a pre-survey was administered in class, which took about 10 minutes. The teacher then gave an introduction to

the first unit, “Solar Energy Science”, which had also been used in other curriculum projects (Sung et al., 2022). Over the next 2-3 days, students worked through the solar energy science unit in a self-directed fashion. The main learning objectives of this section were to describe the Sun’s position using solar elevation and azimuth angles, describe the daily and seasonal changes of solar angles, describe the relationship between the angle of incidence and the energy output of a solar panel, explain the optimal tilt angles of a solar panel that maximizes the energy output in each season and in a year. Each activity followed the Predict-Observe-Explain framework (White and Gunstone, 2014). For example, students would first predict the best tilt angle for fall, then conduct an investigation in Aladdin, where they compared the simulated daily energy output of solar panels with different tilt angles. Finally, they were asked to explain this result using the solar energy science concepts they learned earlier, such as the solar elevation angle and projection effect. At the end of the unit, students completed a challenge called “Optimize It!”, where they needed to find the best position and angle to place a single solar panel in a yard surrounded by trees, such that the panel would generate the most yearly output.

After students completed the Solar Energy Science unit, they continued to the Solar Farm Design unit, where they were tasked with designing a solar farm that would generate the most yearly profit for their town. Here, the main learning objectives were to evaluate a design solution using the given criteria and constraints, collect evidence of the design performance using computer simulations, improve the design performance through iterations, create a design that meets the given criteria and constraints, and explain the choice of design variables using scientific principles. Students were first introduced to the design criteria, constraints, and variables of a solar farm. The design process began with the students evaluating an existing solar farm design with suboptimal performance: a negative yearly profit. Students brainstormed how they could change the three design variables - tilt angle, row width (RW), and inter-row spacing (IRS) - to improve the performance and were asked to document their reasoning as well. After choosing one design variable to change and specifying what value to change it to,

students input the new design variable into Aladdin's layout wizard, which would automatically update the solar farm design layout based on the specified variables. Students were directed to save their new design as a new file on the cloud storage as a method of design documentation and version control. Students then compared the performance of their new design with that of the previous iteration by analyzing the yearly energy output and profit, and they reflected on their learning during this iteration. An example iteration was provided on the design instruction to help explain the design process and clarify the expectation of student responses. The teacher also demonstrated how to go through a design iteration in Aladdin on the projector screen. After that, students were given at least one class period of time to create solar farm designs of their own, and they were directed to document their full design process, including their design variables, performances, and reflections, in a pre-formatted design journal. Students were supposed to work on their own designs, but they were encouraged to discuss with their classmates. While the students were iterating, the teacher was instructed to circulate the classroom, check on student progress, and answer questions.

After students had had a chance to create at least 2-3 designs, they were introduced to the AITA. Using a genetic algorithm with preset parameters, the AITA would use the current student design as the starting point, generate new designs by mutating the current design, improve its strategy by learning from the analysis result of each iteration, and evaluate a total of 50 new designs over 5 generations. While the AITA iterated through different designs, the students could view an animation of how one design changed into the next design and evolved into the final design over time. The final best design (including the design variables and the performance) would be reported as an interactive graph alongside all previous iterations (Figure 2). Students would then answer a series of reflection questions on a worksheet to document their reaction to AI's design and use it as feedback to think about how they can further improve their design. They then had until the end of the implementation to keep iterating (Figure 3). The AI's design also doubled as a formative assessment of student design because it was not

guaranteed that the AITA could find a better design: If the student design was already close to the local optimum, then the AITA would be less likely to find a better design or improve by any significant amount. A post-survey was administered on the last day of the implementation.

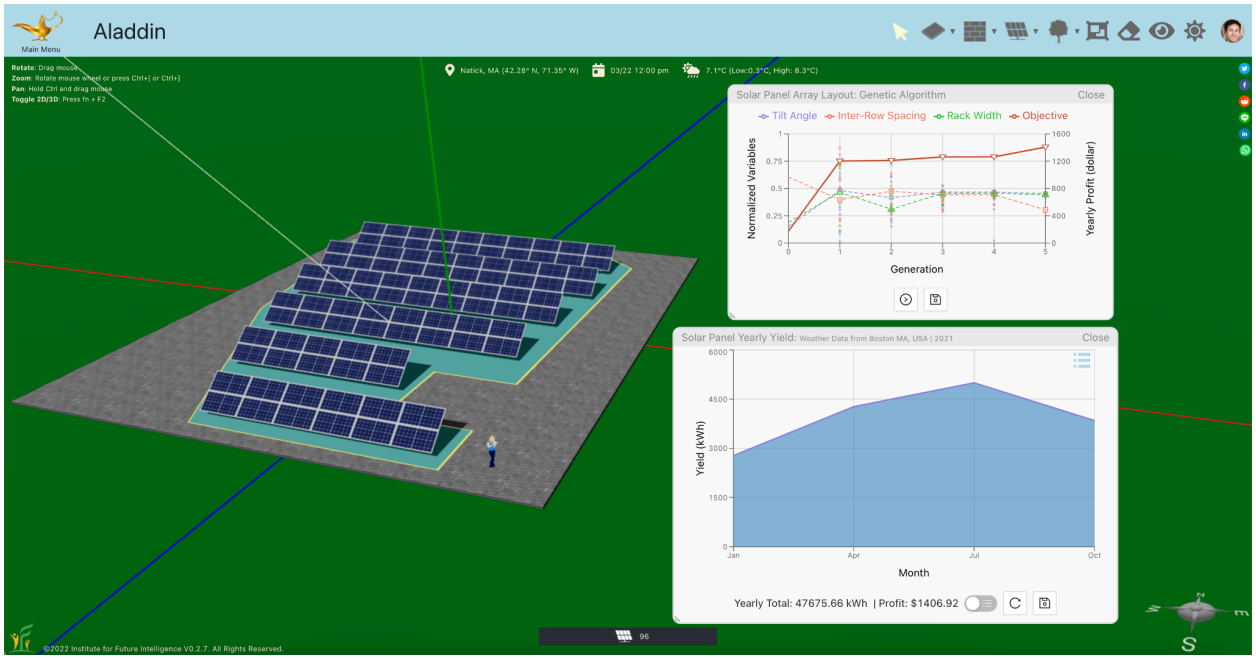


Figure 2. A sample screenshot of an AI-generated solar farm model in Aladdin. The top window shows the evolution of three design variables (tilt, RW, and IRS) and one objective (yearly profit) over multiple iterations. The bottom window shows the yearly yield analysis of the current design.

Figure 3. Classroom photos taken during the implementation: (a) A student from school 1 was inputting new design variables into Aladdin's layout wizard to test a new solar farm design, after reviewing the AITA's design and analyzing its yearly energy output and profit. (b) The science teacher from school 2 was giving feedback on one student's solar farm design.

EVALUATION

The goals of the evaluation were to address the following research questions: 1) To what extent and in what ways does the Solar Farm Design curriculum affect students' learning outcomes? 2) To what extent and in what ways does the use of AITAs affect students' learning outcomes? 3) To what extent and in what ways does the use of AITAs affect students' perception of AI?

Methods

The main types of student data included the following sources:

- Pre- and post-test performance: Students were asked to fill out the same survey before and after the implementation of the solar farm design project. The survey was developed by the authors and consisted of the following components:
    - Multiple choice questions that assess student understanding of the following solar energy science concepts:
        - Daily changes of solar angles
        - Seasonal changes of solar angles
        - Projection effect
        - Optimal solar panel tilt angle
    - An engineering design prior knowledge self-assessment adapted from Carnegie Mellon University (n.d.)
    - An engineering design self-efficacy survey adapted from Carberry et al. (2010)
    - An open-ended question asking students to name at least three important components of the engineering design process. Responses were coded by the researcher (R.J.).
    - (Post-test only) Likert scale questions about student perception of the AITA, adapted from Kim et al. (2020).
    - (Post-test only) Two open-ended questions about what the students enjoyed and would have changed about the curriculum. Responses were coded by the researcher (R.J.).
- AI feedback worksheet: After receiving feedback from the AITA, students were asked to answer a series of reflection questions and given another opportunity to improve their solar farm design. They were instructed to document their pre-AI, AI, and post-AI designs on this worksheet.
- Student interviews: Three interviewers (A.B., I.L., & R.J.) conducted 15- to 20-minute semi-structured interviews after the project implementation. In both schools, students

were selected by the teacher based on availability and interest and covered different levels of engagement and performance. The lead interviewer was determined by availability. All three interviewers followed the same interview protocol developed by one researcher (R.J.), but the exact questions asked varied for each student based on their progress and time availability. Interviews were transcribed by three researchers (A.B., I.L., & R.J.) and analyzed through inductive thematic analysis (Braun & Clarke, 2006) by one researcher (R.J.). The initial open codes were generated after a thorough reading of the transcribed student responses to all interview questions. In a second round of focused coding, only responses containing open codes related to AI or feedback were reviewed, and emergent themes were identified from the final codes and refined into sub-themes.

In addition, the following types of supporting data were collected and used to corroborate the main data:

- Design artifacts: During the solar farm design project, students were instructed to save their solar farm models on the Aladdin cloud storage. When present, these files were used to validate the design documentations on the AI worksheets.
- Student activity log data: Every student action in Aladdin (such as turning on the heliodon, changing the tilt angle of a solar panel, and simulating the yearly energy output) was automatically logged and stored in a database. This log data was consulted when the design documentations on the AI worksheets were incomplete or contained inconsistencies. Note that the data logger in Aladdin was only enabled during the implementation for research purposes, and it is currently disabled for regular users.
- Design journals: During the solar farm design portion of the project, students were asked to document each design iteration in a design journal and answer a series of reflection questions about each iteration. The journals were also used to validate the design documentations on the AI worksheets.

- Teacher feedback: After each class period, the teachers in both schools gave verbal descriptions of their observations in the classroom. An informal interview was also conducted after the project implementation.
- Observation notes: In both schools, the teachers set up additional cameras that were connected to a video conference during the implementation, so researchers could observe and take notes on the classroom dynamics and student engagement.

Compliant with IRB requirements, student assent forms and parent consent forms were distributed prior to the study, and data was only collected from assenting students. Each student was assigned an anonymous ID in the following format: C_P1S1 or M_P1S1. The prefix indicates the school (C for school 1, M for school 2), and P1S1 stands for “period 1, student 1”. All students were only referred to as their anonymous IDs in subsequent data analysis.

RESULTS

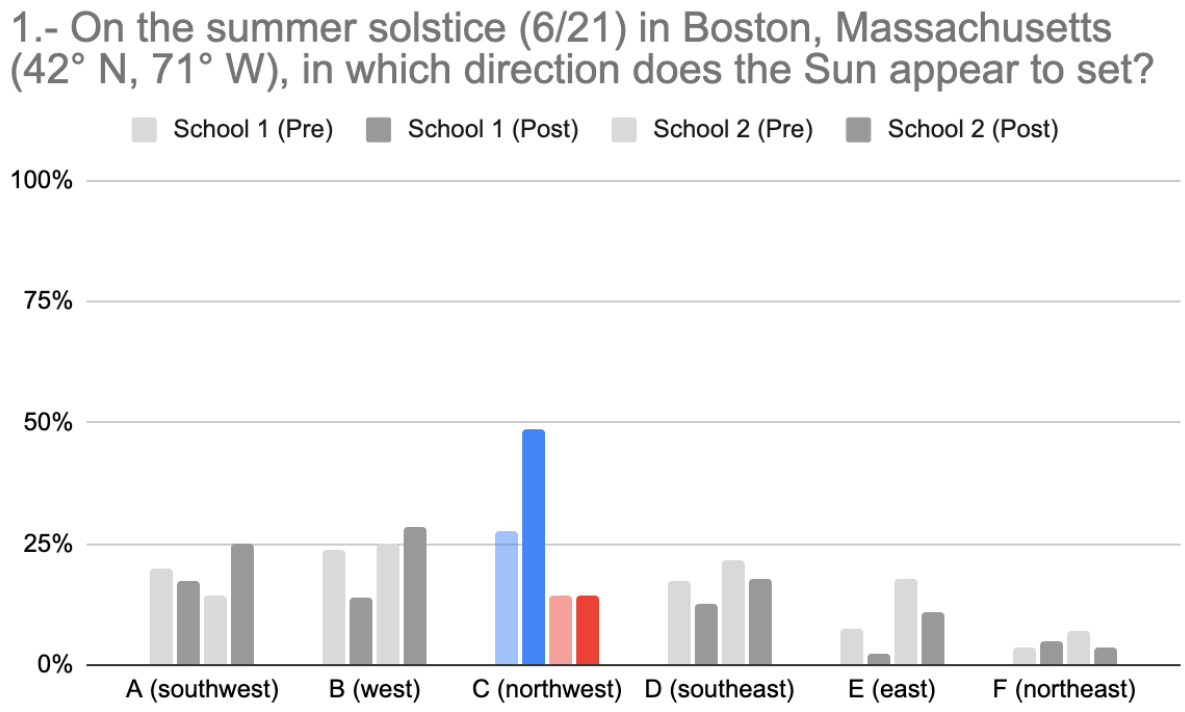
Table 2 showed the availability of different types of student data. Unless otherwise stated, the pre- and post-test data referenced below included only data from 80 students in school 1 and 28 students in school 2 who submitted both tests.

	School 1	School 2
Pre-test	97 / 111	31 / 31
Post-test	87 / 111	28 / 31
Both pre- and post-tests	80 / 111	28 / 31
Design documentation	41 / 111	16 / 31
AI worksheet	25 / 111	1 / 31
Interview	5 / 111	10 / 31

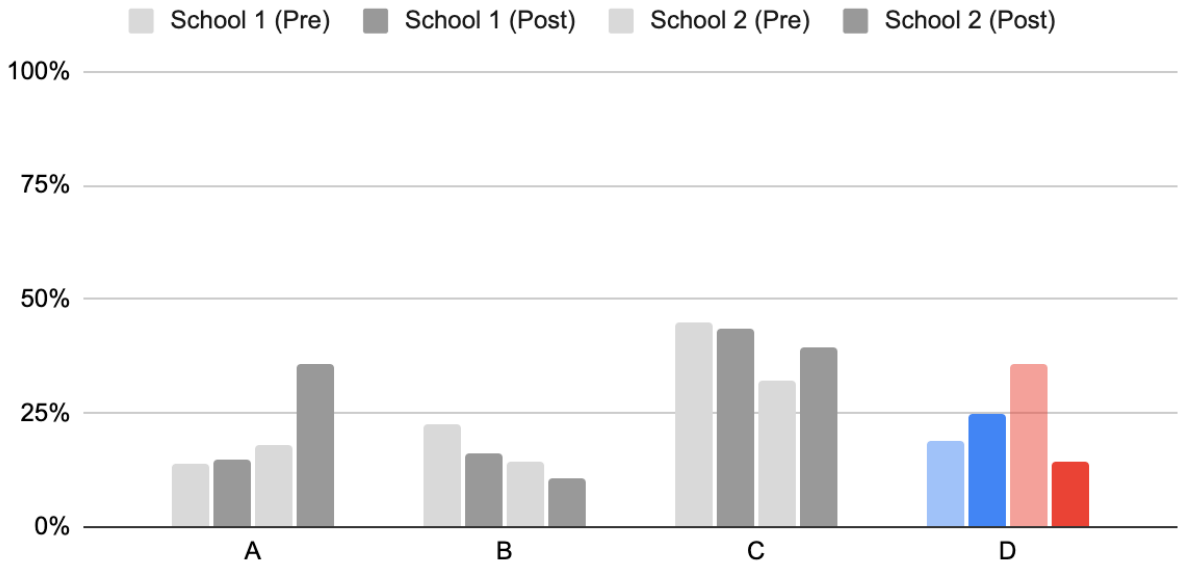
Table 2. A breakdown of student data availability from different sources.

Pre-and post-tests

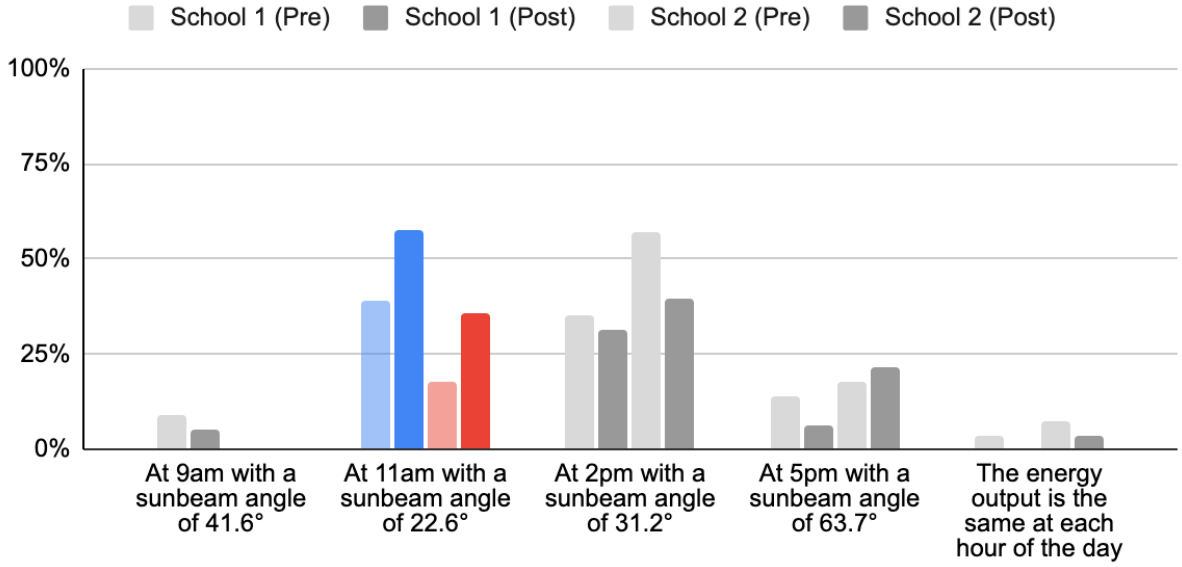
Figure 4a-4d showed the student responses to 4 multiple-choice questions that assess solar energy science knowledge. In school 1, the mean score increased in all 4 questions (Figure 4a-4d). In school 2, the mean score increased only in question 3 (Figure 4c), remained the same in question 1 (Figure 4a), and decreased in question 2 and 4 (Figure 4b & 4d). A paired one-tailed t-test using the aggregate results from the pre-test (M_1 = 1.15, SD_1 = 0.94; M_2 = 1.04, SD_2 = 0.74) and post-test (M_1 = 1.8, SD_1 = 1.08; M_2 = 0.89, SD_2 = 0.74) indicated that the solar farm design curriculum resulted in an improved understanding of solar energy science knowledge in school 1 (p_1 = 0.00002), but not in school 2 (p_2 = 0.25).



2.- The images below show the same tree and its shadow at different times and dates in Boston. Which one is more likely to be at 2pm on the winter solstice (12/21)?



3.- A single solar panel lies flat on the ground in Boston. The date is 6/21 (summer solstice). At which hour of the day would it produce the most energy?



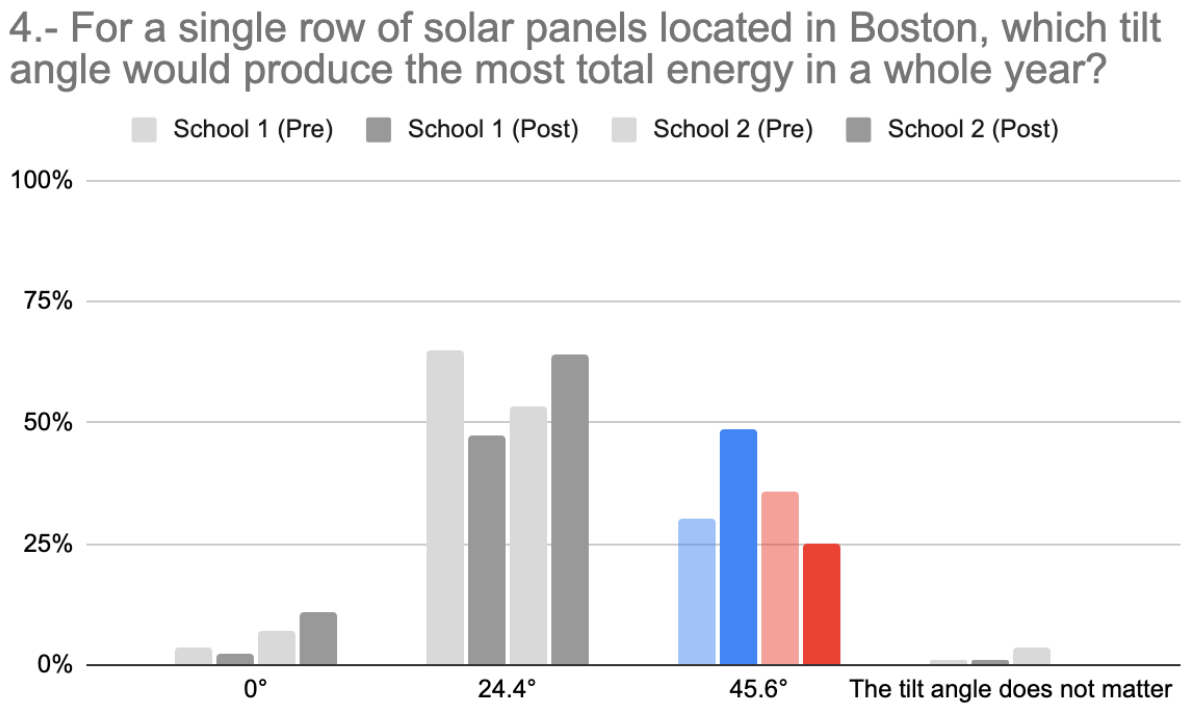


Figure 4. The distribution of student responses to the solar energy science assessment items, before and after completing the solar farm design curriculum. (a) A question about daily changes of solar angles. (b) A question about seasonal changes of solar angles. (c) A question about the projection effect. (d) A question about the optimal solar panel tilt angle.

Figure 5 showed the responses to the multiple-choice self-assessment “How familiar are you with ‘engineering design’?” Before the project, 51.5% of the students in school 1 and 51.6% in school 2 chose level 2 (“have some idea what it is, but don’t know when or how to do it”), followed by level 1 or the lowest level (“I have never heard of it or I have heard of them but don’t know what it is.”) with 35.1% and 35.5% of the votes, respectively. In school 1, over half of the students (54.0%) reported level 4 or the highest level (“I can explain what it is, how to do it, and I have done it”) after the project, followed by level 2 (29.9%). In school 2, the plurality choice remained level 2 (42.9%), while only 21.4% of the students selected level 4.

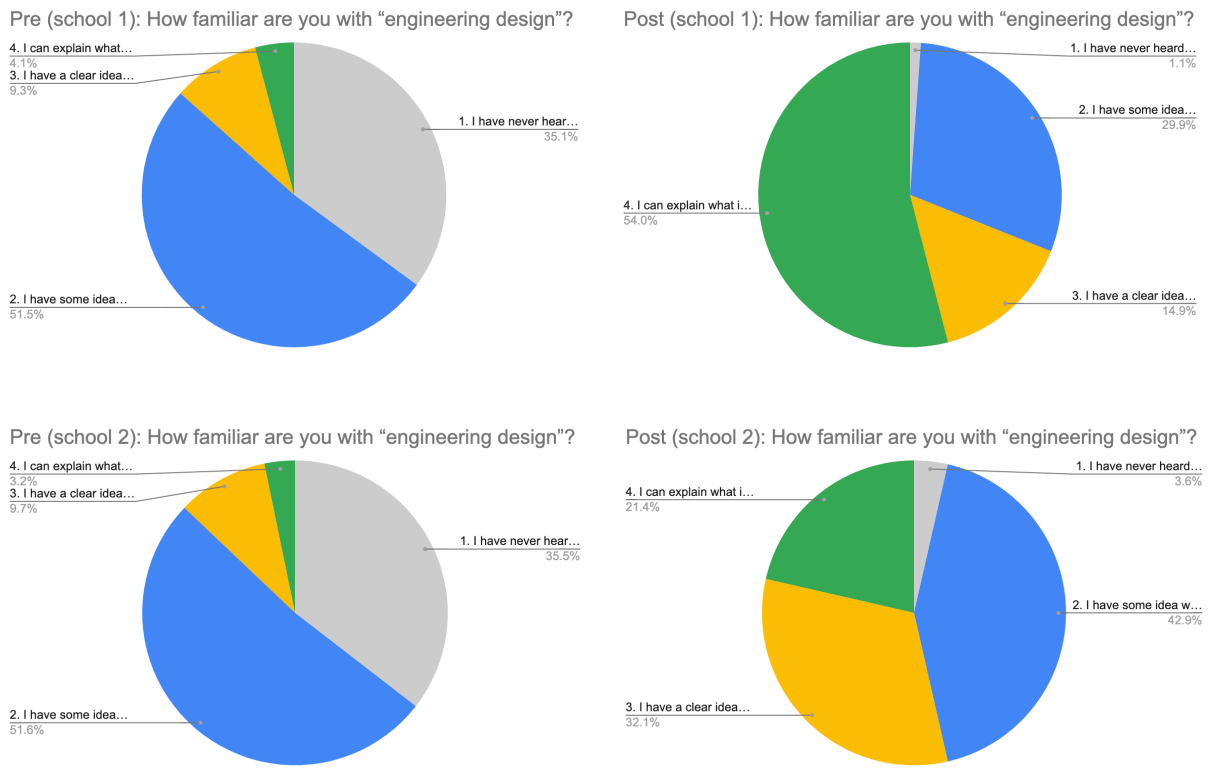


Figure 5. A comparison of students’ self-assessment of their familiarity with engineering design, before and after completing the solar farm design curriculum.

Figure 6 showed the coded responses to the free-response question “ What are the important components of the engineering design process? Name at least three components.” The percentage of non-responses, which included answers like “idk” and “I’m not sure”, decreased by 23% in school 1 and 25% in school 2. For 6 of the 9 categories, the difference between pre- and post-tests were within 10%. In school 1, 25% more students identified “redesign” as an important component of the engineering design process. In school 2, the biggest increase was for “research a design need” (15%), followed by “redesign” and “evaluate” (10%).

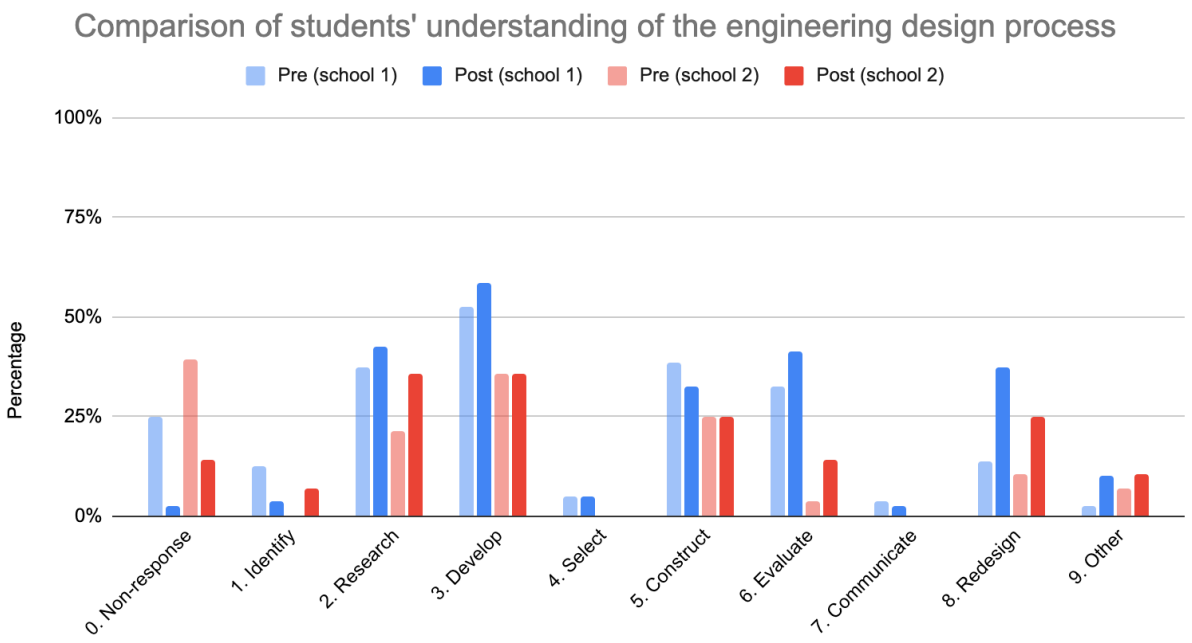


Figure 6. A comparison of students' self-reported understanding of the engineering design process, before and after completing the solar farm design curriculum.

Figure 7 showed students' self-rated confidence with the 8 components of the engineering design process according to the Massachusetts Department of Education (2006). The average ratings were 45 (school 1) and 36 (school 2) before the project and increased to 66 (school 1) and 58 (school 2) after the project.

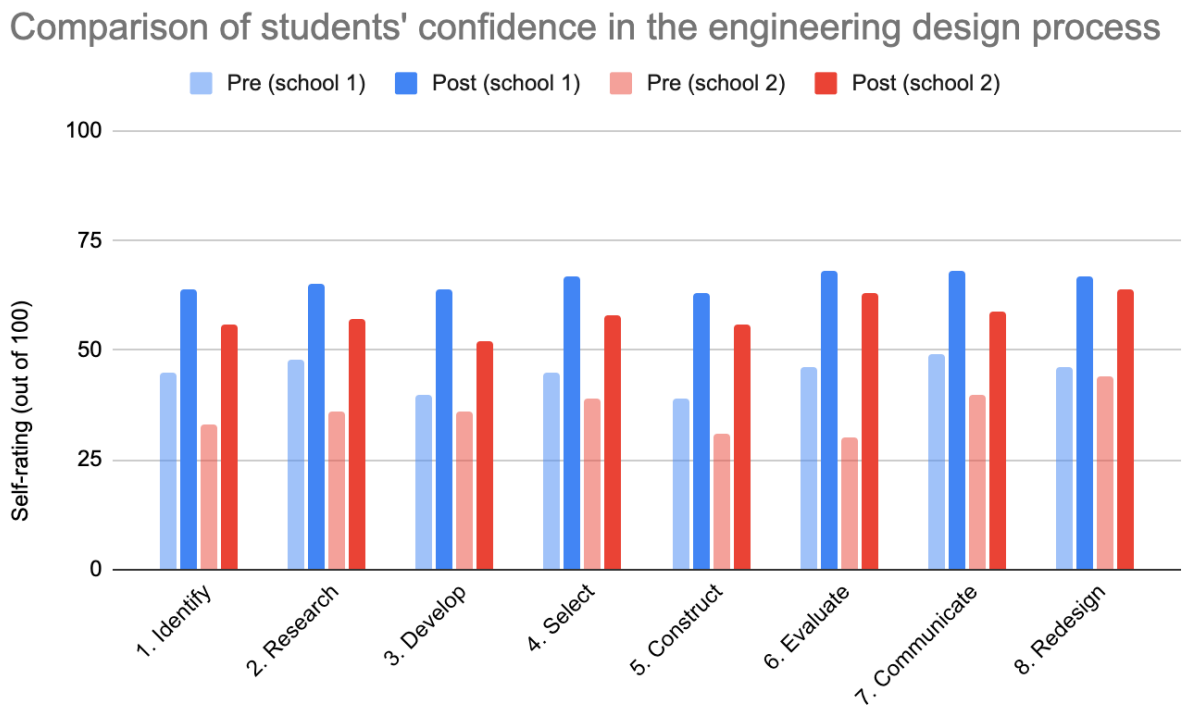


Figure 7. A comparison of students' self-rated confidence in the engineering design process, before and after completing the solar farm design curriculum.

Figure 8 showed students' overall perception of the AITA in the post-survey. The responses were more positive in school 1 (M_1 = 4.38, M_2=4.25) and more spread out in school 2 (SD_1 = 1.53, SD_2 = 1.90).

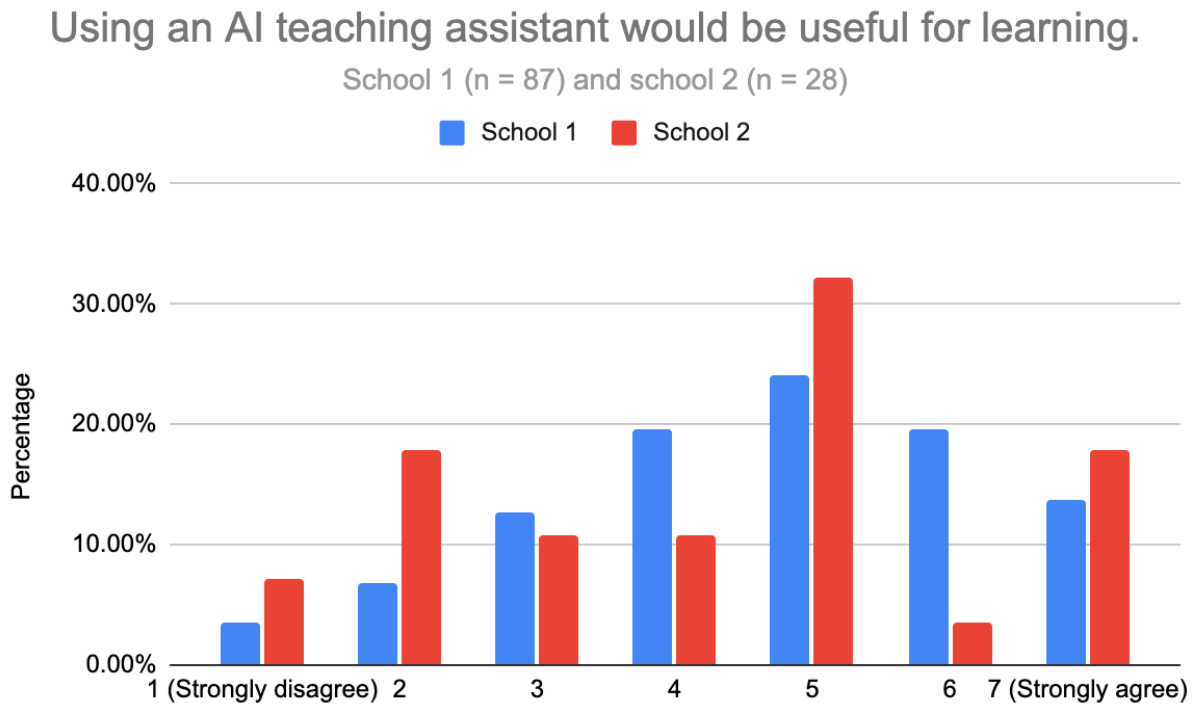


Figure 8. Responses to a 7-point bipolar Likert scale question about student perception of the AITA, with 1 being “strongly disagree” and 7 being “strongly agree”. The remaining numbers were not labeled in the survey.

Engineering design performance

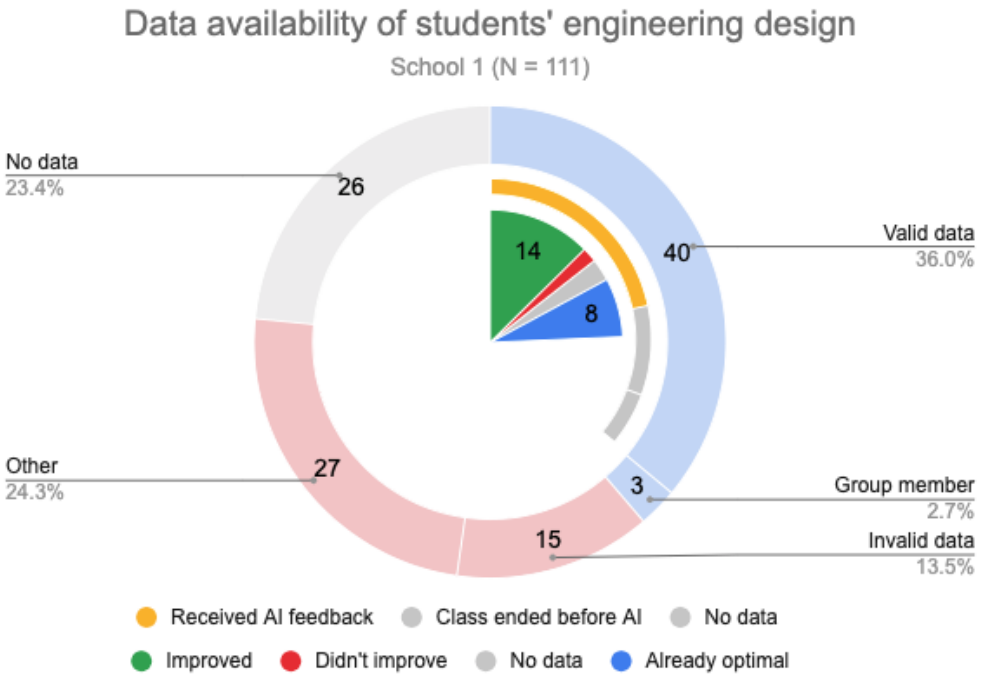
Data availability

Students’ design data (including the pre-AI design, AI design, and post-AI design) were organized from AI worksheets and design journals and validated using the log data. A student’s design data was considered to be complete if it contained all 3 designs. A student’s design data was considered to be valid if there were no inconsistencies among different data sources. A student’s design data was considered to be unique if the student didn’t share their data with anyone else in a group setting.

Figure 9a showed that in school 1, 40 of 111 students left complete, valid, and unique documentation of their solar farm designs before receiving AI feedback. Of the 81 students who

were excluded from further discussions of design performance, 3 students worked together with other students and used their data with permission; 15 documented data that were incomplete or incomparable with other students' data; 27 used other students' data without express permission or documented data that couldn't be validated by other data sources; 26 didn't finish the activity or document enough data. Of the 40 students with valid data, 24 documented the feedback from AI; 10 were from a class that had to end early before the AI activity; 6 didn't document enough data. Of the 24 students that received AI feedback, 14 increased the yearly profit of their final design; 2 didn't find a better design than AI's recommendation; 3 didn't iterate again or document enough data; 5 students already had near-optimal designs. In addition, 3 more students who didn't receive AI feedback also had near-optimal designs. Since the AITA was unlikely to improve a near-optimal design within 1 run (or 50 iterations), students with near-optimal designs were excluded from any analysis or discussion of AI feedback.

Figure 9b showed that in school 2, 16 of 31 students left unique and valid documentation of their solar farm design before receiving AI feedback; 7 documented data that were incomplete or incomparable with other students' data; 8 didn't finish the activity or document enough data. Of the 16 students with valid data, 10 documented the feedback from AI; 6 didn't document enough data. Of the 10 students that received AI feedback, only 1 student increased the yearly profit of their final design; 9 didn't iterate again or document enough data.



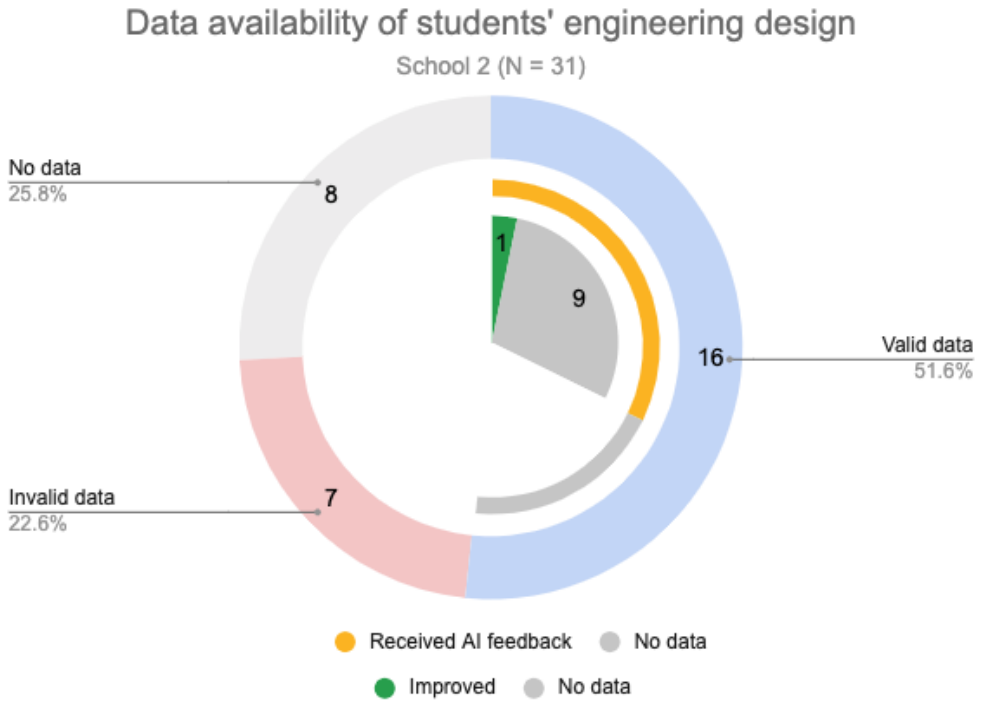


Figure 9. A breakdown of the data availability of students' engineering design. The outer ring showed the sample size of student designs before the AI feedback and reasons why data points were excluded. The middle ring showed the sample size of students who received AI feedback

and reasons for exclusion. The inner pie chart showed students' reactions to AI feedback. (a) Statistics from school 1. (b) Statistics from school 2.

Design space and benchmark performance

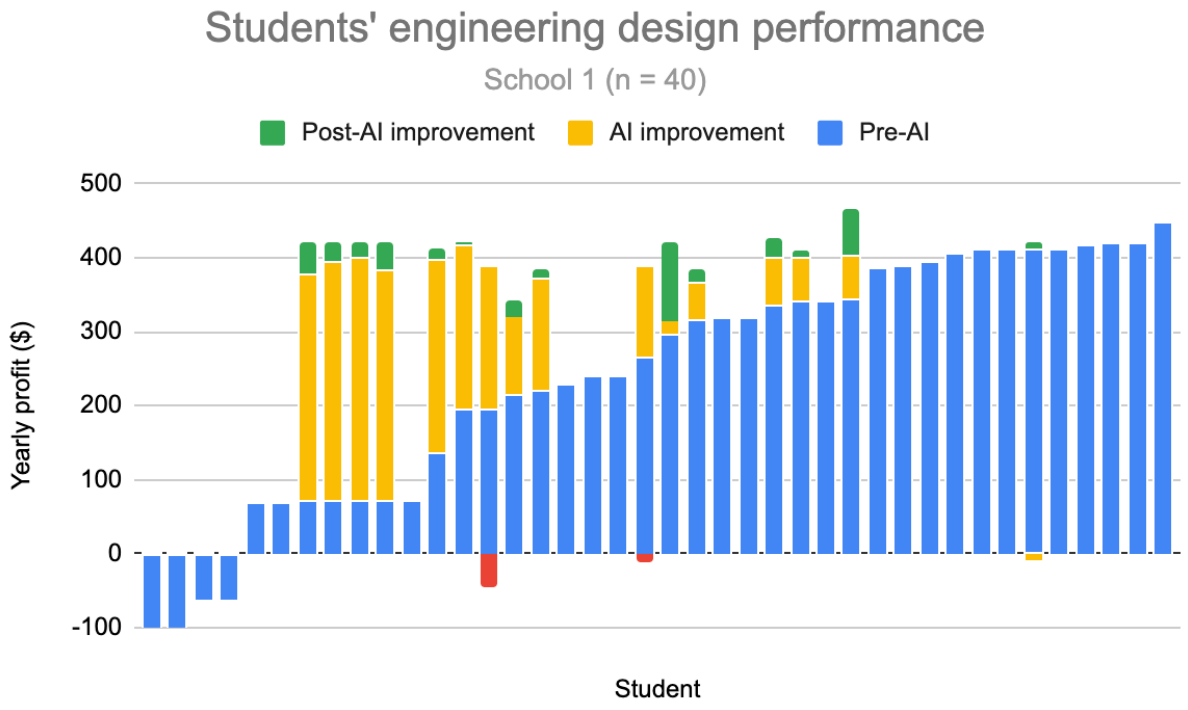
Students' engineering design performance was evaluated using a single metric: The yearly profit of their solar farm design, which equals the revenue (determined by the total energy output of the solar panels) minus the cost (determined by the number of solar panels used). In general, a tilt angle equal to the latitude of the location (around 42º for the two schools) would optimize the energy output per solar panel for an entire year, although the curriculum placed a wind resistance constraint that limited the maximum tilt angle to 35º. The other two design variables, row width and inter-row spacing, were coupled: A larger RW required a larger IRS to avoid inter-row shading. Therefore, an optimal design was one with an optimal tilt angle and a suitable pairing of RW and IRS that fit as many solar panels onto the given plot as possible while minimizing inter-row shading. Solar farm designs would be hereinafter denoted in the following format: (tilt, RW, IRS).

The allowed design space was specified as (0º-35º, 1-6 panels, 2m-10m), which was also set as the search range of the AITA's genetic algorithm. At least three local optima existed within this design space: (35º,1 panel, 2.3m), (35º, 2 panels, 4m), (35º, 3 panels, 7m). When deployed on the given plot in the curriculum, these optimal designs produced a yearly profit of around $420 in school 1 and $517 in school 2. The difference was due to different weather conditions. A student design was considered to be optimal or near-optimal if its yearly profit was greater than $400 in school 1 or $500 in school 2.

Figure 10a showed that in school 1, 36 out of 40 students were able to make a profit on their solar farm design before receiving AI feedback (median = $282, SD = 362.40). 16 of the 36 students received AI feedback that improved the yearly profit of their designs (median = $392,

SD = 28.92). 14 students further improved their profit after receiving AI feedback (median = $420, SD = 32.50), whereas 2 students didn't.

Figure 10b showed that in school 2, 11 out of 16 students were able to make a profit on their solar farm design before receiving AI feedback (median = $266, SD = 561.00). Only 1 student recorded the AI feedback, which failed to improve the yearly profit due to a technical issue in the early version of the Aladdin software. However, the student improved their own design after receiving AI feedback, nonetheless.



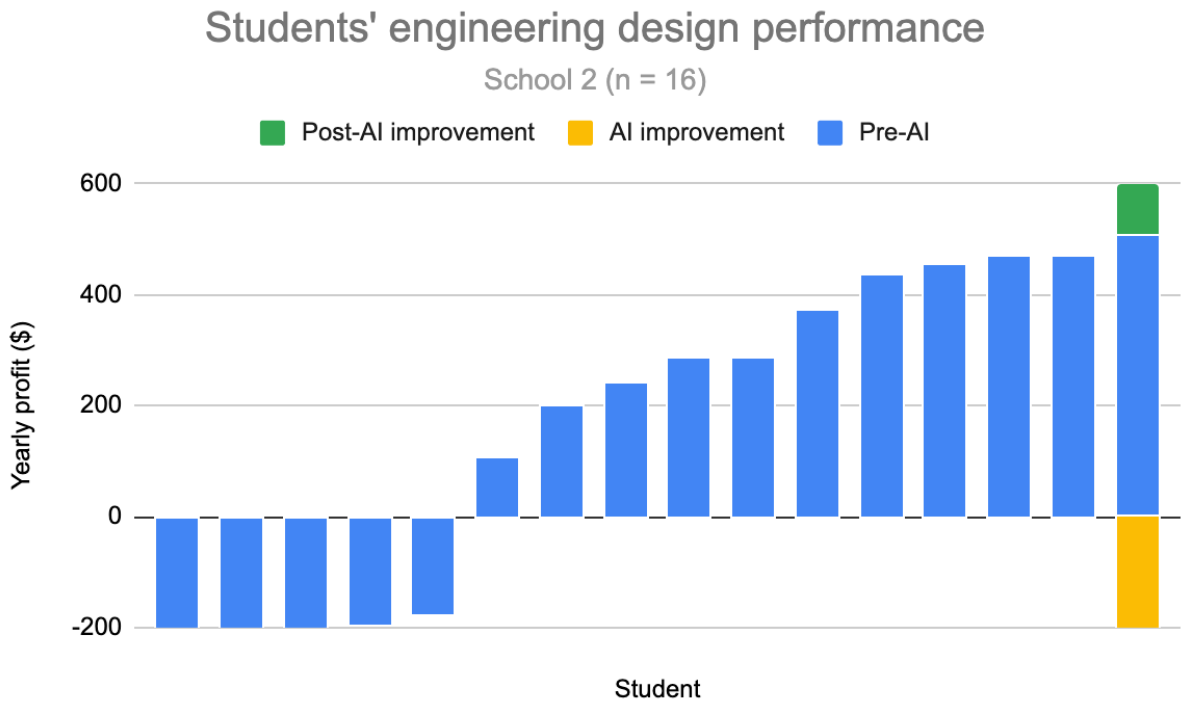


Figure 10. A comparison of students' engineering design performance (measured by the yearly profit of their final solar farm design), before receiving AI feedback (blue), the improvement made by AI's recommended design (yellow), and the improvement made by students' design after they received AI feedback (green). The negative y-axis was truncated to save space. Red columns below zero indicated the difference in yearly profit between students' post-AI design and AI's design when a student couldn't further improve AI's design. Yellow columns below zero indicated the difference between students' pre-AI design and AI's design, when the AI couldn't improve a student's design, which were rare faulty behaviors from AI. (a) Statistics from school 1. (b) Statistics from school 2.

Students' reactions to AI feedback

Due to the lack of AI feedback data from school 2, only AI feedback for 16 students in school 1 was manually categorized by the researcher (R.J.) by comparing AI's designs with

students' pre-AI designs (Figure 11). The tilt angle was analyzed separately due to it being relatively independent from the other two design variables. To better illustrate the common themes in AI feedback and student reactions, a few case studies were presented below.

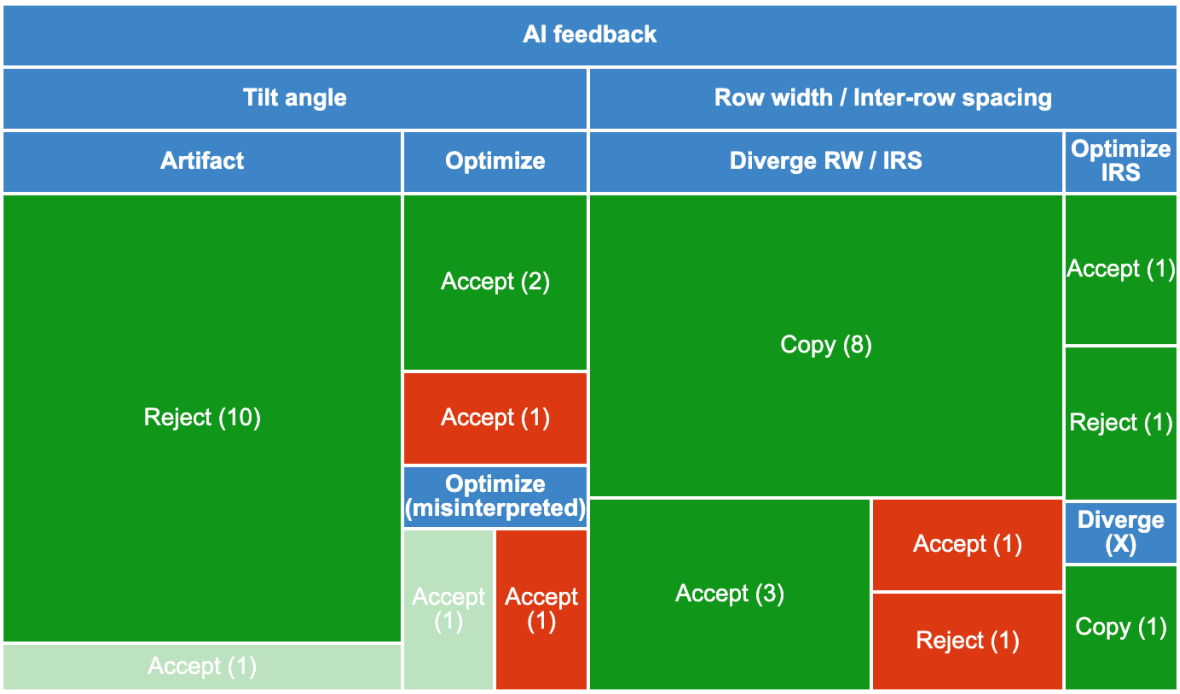


Figure 11. A vertical treemap chart showing students' reactions to different AI feedback. Blue tiles represented different categories of AI feedback. Dark green tiles represented student reactions that led to further improvement of the yearly profit in the post-AI design. Red tiles represented reactions that didn't lead to any improvement. Light green tiles represented reactions that wouldn't have led to any improvement per se but still did due to optimization of other design variables.

C_P5S29: Best of both worlds

The first case study, C_P5S29, illustrated the most common AI feedback and the most common reaction. The student's pre-AI design was (35º, 2 panels, 12m) and made a yearly profit of $220. AI's recommended design was (33.87º, 3 panels, 8m) and made $374 of profit.

The student was “not really” surprised by AI’s tilt angle because he “knew the optimal range is between 30º-60º”, but he rejected AI’s change and kept his original tilt angle in his final design. This common AI feedback to decrease the tilt angle, received by 11 students total, was considered an artifact of the software algorithm because the AI search range for tilt angle was capped at 35º, and AI could only output floating point numbers that were bound to be less than 35º. 10 of 11 students made the correct decision to reject this feedback.

On the other hand, he was “very much” surprised by the other changes because “minor changes created such drastic differences”. He ended up copying AI’s exact RW and IRS in his final design, (35º, 3 panels, 8m), which made $383 of profit. When AI used a different RW than the student did, it was interpreted as feedback to diverge more in the design space. 14 students received this feedback, and 9 students copied AI’s exact same values in their final designs.

On rare occasions, AI was able to inspire students to diverge even when its own attempt failed. For example, C_P5S16’s pre-AI design was (35º, 2 panels, 4m) and made $412 of profit. When they saw AI’s design of (33.21º, 3 panels, 6.89m), they were not surprised by the wider rows, writing: “In my initial testing, I found that 3 panel width was effective in creating a large amount of profit. Only after much more testing did I find a much more effective 2 panel width design. So, it does not surprise me that the AI found and stuck with the 3 panel width idea.” Their design journal, which had over 10 iterations, showed that they had already tried (35º, 3 panels, 6.5m). Their post-AI design of (35º, 3 panels, 6.89m) combined AI’s wider rows and their original tilt angle and made $421 of profit.

In total, 7 students responded in the same way to AI feedback: They kept their original tilt angle when AI used a lower one, copied AI’s RW and IRS, which were different from their original ones, and improved their profit in the end by combining the best of both worlds.

C_P4S30: Go the extra mile

In the case of C_P4S30, The student’s pre-AI design was (35º, 2 panels, 7m) and made a yearly profit of $297 (Table 3). The AITA recommended the design (33.87º, 3 panels, 8m) and

increased the yearly profit to $314.44. The student also reported being “surprised” by AI’s changes to the RW because “an increase in the [row] width by 1 panel resulted in such a big difference in energy output”. They went on to explain the rationale of this surprise: “If there are more solar panels per row, then the yearly energy output will increase because the additional panels receiving sunlight contribute to the amount absorbed.”

Design variables	**Your final design**	**AI’s final design**	**Are you surprised by AI’s changes? (Not really / Somewhat / Very much)**
Tilt angle (º)	35	27.47	Somewhat
Row width (panels)	2	3	Very much
Inter-row spacing (m)	7	7.55	Not Really
Design Criteria	**Your final design**	**AI’s final design**	**How did AI’s changes affect this criterion? (increased / decreased / kept the same)**
Number of solar panels	66	96	Increased
Yearly energy output (kWh)	30094.77	43305.87	Increased
Yearly profit ($)	296.69	314.47	Increased

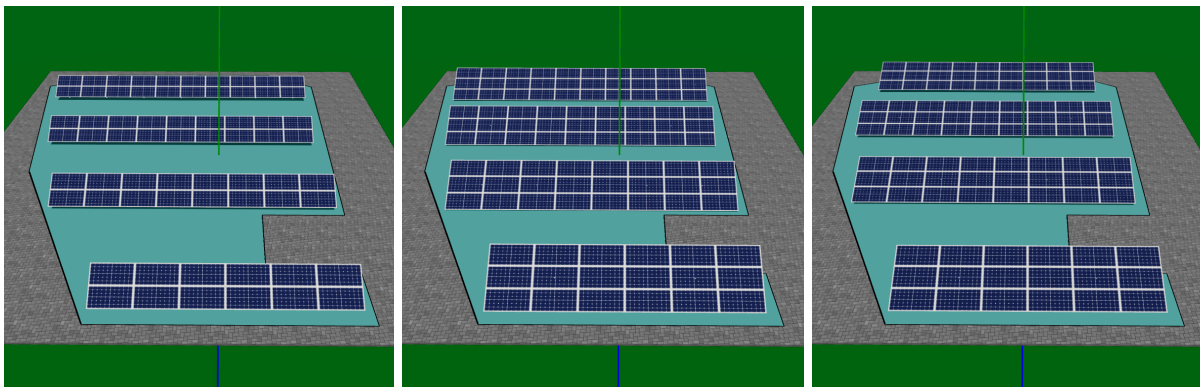
Table 3. A partial reproduction of C_P4S30’s AI feedback worksheet showed their reaction to the AITA.

In their reflection, the student agreed with the change in RW, while acknowledging that there could be more optimal tilt angles. They didn’t agree with the change in IRS, citing the decrease in solar panels that would ensue. The student then documented 3 design attempts (Table 4 and Figure 12), the best of which made $420.62 of profit by integrating AI’s new RW of 3 panels into their original design to create the final design (35º, 3 panels, 7m).

Design variables	**Your design**	**AI’s final design**	**Your new design 1**	**Your new design 2**	**Your new design 3**
Tilt angle (º)	35	27.47	35	30	30
Row width (panels)	2	3	3	3	2
Inter-row spacing (m)	7	7.55	7	8	5
[redacted]					

Number of solar panels	66	96	99	90	90
Yearly energy output (kWh)	30094.77	43305.87	45044.48	40803.92	40790.26
Yearly Profit ($)	296.69	314.47	420.62	345.98	342.57

Table 4. A partial reproduction of C_P4S30's AI feedback worksheet showed the documented evolution of their solar farm design, before and after receiving feedback from the AITA.



(a) (b) (c)

Figure 12. Side-by-side comparison of three solar farm designs. (a) C_P4S30's best design before receiving AI feedback. (b) The AITA's design, which used wider rows and a lower tilt angle than (a). (c) C_P4S30's best design after receiving AI feedback, which used the same wider rows as in (b) but reverted to the higher tilt angle used in (a).

Instead of directly copying AI's design variables, C_P4S30 accepted AI's feedback to diverge and diverged further in the design space by combining AI's wider rows with their own tighter IRS. 2 other students reacted similarly by going the extra mile. C_P6S18's pre-AI design was (32º, 2 panels, 6m). When they saw AI's design of (30.66º, 3 panels, 7.37m), they pivoted to wider rows and further increased the IRS, creating the final design (35º, 3 panels, 8m). On the other hand, C_P620's pre-AI design was (35º, 3 panels, 10m). When they saw AI's design of (33.76º, 1 panel, 2.33m), they kept their wider rows but took inspiration to decrease the IRS, creating the final design (33.76º, 3 panels, 6m).

C_P4S19: You may already know this

When AI did use a higher tilt angle, which was interpreted as feedback to optimize the tilt angle, 2 students improved their design by accepting the feedback and further increasing the tilt angle. For example, when AI changed the tilt angle of C_P4S19's design from 30º to 34.5º, she further increased it to 35º. Reflecting on this decision, she wrote: "We noted that any degree closer to 35 is more ideal. So even that slight change of half a degree can make a dollar or more difference over the course of a year." Another student C_P4S14, who didn't document their post-AI design, was "surprised" to see the AITA increase the tilt angle of their design from 25º to 27º and wrote: "I say this because I did not think about further tilting the panels as I felt I had reached the sweet spot. Now, I understand that the further tilt accommodates the middle seasons, fall and spring, much more appropriately and therefore produces more energy."

C_P5S5: Unintended consequences

When AI used design variables that were similar to what students used, the feedback became more difficult to interpret and led to some misconceptions. For example, when C_P5S5 compared AI's tilt angle of 30.09º with their own tilt angle of 30º, they wrote: "I wasn't surprised because I feel like 30 degrees is the best tilt angle." They did try other tilt angles after receiving AI feedback, but when none of their attempts improved AI's design, they concluded that "the [30º] angle is perfect". Similarly, when C_P5S20 compared AI's tilt angle of 32.74º with their own tilt angle of 30º, they wrote: "I am not surprised because I know that the tilt angle is most efficient in the low thirties due to the seasons." When their best post-AI design, which used a slightly higher tilt angle of 31º, turned out to be better than AI's, they still concluded: "If the solar panel tilt is closer to 30 degrees, the amount of total energy increases."

Another factor that contributed to these misinterpretations was the lack of controlled experiments. Both C_P5S5 and C_P5S20 changed multiple design variables at a time in their post-AI iterations and didn't do any controlled experiments to isolate the effect of tilt angle, leading to their misconception that 30º was the best tilt angle.

C_P5S2: No pain, no gain

C_P5S2 was the other student that didn't improve AI's design. Their pre-AI design was (30º, 1 panels, 3m), and when AI recommended the design (32º, 3 panels, 7m), they rejected the feedback to diverge to 3-panel rows and kept using 1-panel rows. That decision could still have led to success, since the student was already near the local optimum (35º,1 panel, 2.3m), which would have made over $400 of profit, but the student stopped short at (34º,1 panel, 2.5m), which made $10.5 less than AI did. Similarly, C_P5S5 also could have improved AI's design, had they learned to change only one variable at a time and iterated a few more times.

Observations and Interviews

During the implementation, students from both schools had access to three types of feedback: In addition to the AI feedback towards the end of the project and occasional verbal feedback from the teacher throughout the project, it was observed that students would frequently (though not always) discuss with their peers during their independent work time. Table 5 showed the relevant interview questions that were centered around how students experienced different types of feedback, and Table 6 showed the results of the thematic analysis of the interview transcripts.

Relevant interview questions	Count
1. What part of the design process was the MOST engaging for you?	13
2. What part of the design process was the LEAST engaging for you?	13
3. What kind of feedback did you receive on your design, if any?	12
4. How would you compare receiving design feedback from a teacher, a classmate, and AI?	7
5. If you can change one thing about how AI gives you design feedback, what would you change? Why?	7
6. If you can receive feedback on anything, what kind of feedback do you	10

think will help you the most with improving your design?	

Table 5. A list of interview questions related to design feedback.

Themes	Sub-themes	Definition	Example quote	Count
Values in feedback	Visualization	The feedback recipient can visualize designs and design changes in different representations.	C_P5S25: “I'm a super visual learner. So, being able to see the whole program work itself through and see the pictures behind it and be like ‘oh, you can see it's at this angle’ or ‘this is where the shading part is’ and ‘this is the best part to put this.’ That's what I really like about it.”	2
	Authority	The feedback giver is deemed trustworthy due to their perceived experience in the subject matter.	C_P5S25: “The teacher won’t know everything. The teacher can take a guess, but AI probably knows the best design because you can see it run through 100 designs”.	2
	Empathy	The feedback giver can express emotions and relate to shared experiences.	M_P5S3: “Mr. [teacher] isn't a robot. He can show emotion, so he was smiling, where[as] the bot was just giving me information of what I could do to do better”.	2
	Usability	The feedback recipient can receive, understand, and act on the feedback easily.	C_P4S19: “It [AI] was like a lot of things that you would press and steps you gotta go through. That was confusing.”	4
Types of feedback	Comparative	The feedback shows the recipient other people’s results.	M_P5S10: “It was nice to actually compare [to] someone’s numbers.”	5
	Directive	The feedback provides explicit steps the recipient needs to take to improve.	M_P5S16: “"If you move it at a slanted angle and push it backwards a little	4

			bit, you might have better results.”	
	Facilitative	The feedback engages the recipient in independent thinking and sensemaking.	M_P5S10: “Maybe it [could] explain why they thought more panels were better, or why [changing] 10 to 12 makes a huge difference.”	2
Affect from feedback	Surprise	The feedback leads the recipient to unexpected conclusions.	M_P6S9: “I was surprised by it, because obviously [it was] a lot better than mine.”	1
	Challenge	The feedback recipient views the feedback as a challenge to win a competition.	M_P5S10: “We need to try and beat it because it made it a challenge.”’	3
Additional effects of Feedback	Intention	The feedback helps the recipient figure out the goals and next steps.	M_P6S9: “[I] kinda just… thought, ‘you know I’ll just put in some random number and see what comes out’. But as I [saw AI’s design, I] realized that maybe I didn’t need to have something lower or something higher… The AI knew that right off the bat.”	3
	Confirmation	The feedback confirms whether certain decisions are good/bad for the recipient.	C_P5S29: “I just learned that my assumptions about the tilt angle were correct.”	1

Table 6. A list of themes and subthemes from the interviews.

Students described various qualities they valued when receiving feedback on their designs. In general, AI feedback excelled at providing a visual aspect and earning trust with its computational power. For example, when asked from whom they prefer to receive feedback, C_P5S29 picked AI feedback because AI “can do more calculations than the teacher can or would want to”. On the other hand, AI lacked the empathy that a human teacher or a peer could

provide. When M_P5S16 explained why they preferred peer feedback, they said, “because they [my classmates] are in the same kind of position as me. Because they're also experimenting with it as we do it together, and it kind of helps. Sometimes we might realize something and be like, ‘Oh, yeah, make sure to do that.’” There were also 4 accounts of students complaining about the software crashing or being difficult to navigate, which hindered its usefulness.

Students reacted favorably to the type of comparative feedback provided by the AITA. M_P6S9 found the AI feedback “straightforward” because “it was like seeing, ‘Oh, I did this one way, and the AI did this another way’. Even for a student who didn’t interact with AI like C_P3S12, they liked to “look at numbers” and preferred to receive feedback by seeing “the numbers compared to another design, or see the output compared to my output”, so that “the numbers would make more sense to me.” Students also offered a variety of ideas for how the AITA could give different types of feedback. M_P5S16 would like “more specific” feedback, such as "it's a little bit too slightly to the left". C_P3S28, who didn’t interact with AI, also wanted directive feedback such as “Oh, other students had success changing the row width. Maybe you should try doing that.”

In addition to design improvements, some students reported that AI helped them figure out what to do in the design process. For example, M_P5S10 described how the AITA was “where it all clicked, and it made sense on what everything was doing”. He explained that “before we started making our own [solar farm], I was just kind of pressing buttons and watching things change… I didn’t know really what to do… And then, once we did our own [solar farm], the directions actually made me change them and I watched them change, like the visual aspect.”

Finally, there was further evidence that the presence of an AITA created affective responses among students. Students reported being surprised by the AITA, which was also observed in the student reflections from the AI worksheet. M_P5S10 also claimed the most engaging part of the design process was “when … we built one [solar farm], and then we

compared [ours] to the AI." Engagement didn't guarantee a positive experience throughout, but there could be payoff at the end. According to M_P5S5, "I was hearing how a little bit of people were struggling, but then they were happy when they beat the AI."

Student interview data showed no dominant source of feedback (teacher, peer or AI) that was preferred by the majority. Instead, students provided different reasons for preferring each source of feedback. For example, M_P6S7 elaborated that they "might take in my fellow classmates if they were doing really good, and they were improving multiple times each time." It's worth noting that some students also expressed a preference for multiple sources of feedback. For example, M_P5S16 commented "I would compare it [the feedback] and see if a lot of the things that they mentioned matched up with one another… If a student and a teacher said that the positioning was a little bit awkward, because it's 2 [people], then it has more of a stronger standpoint."

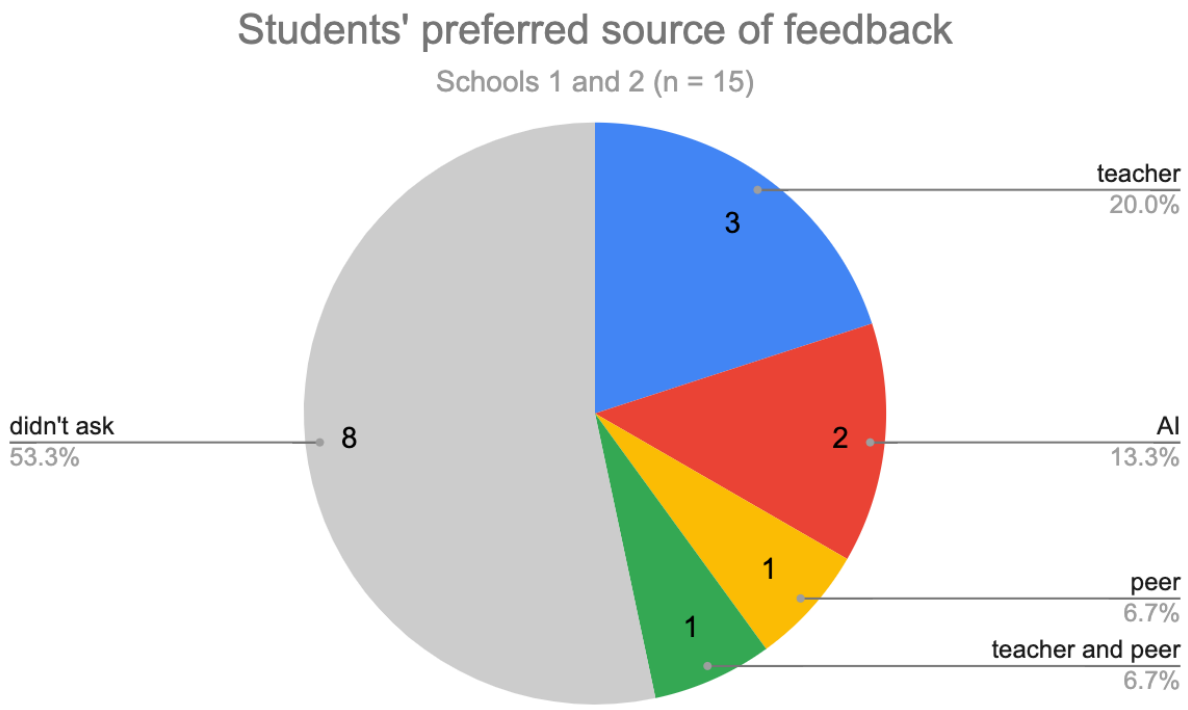


Figure 13. A breakdown of the students' preferred source of feedback.

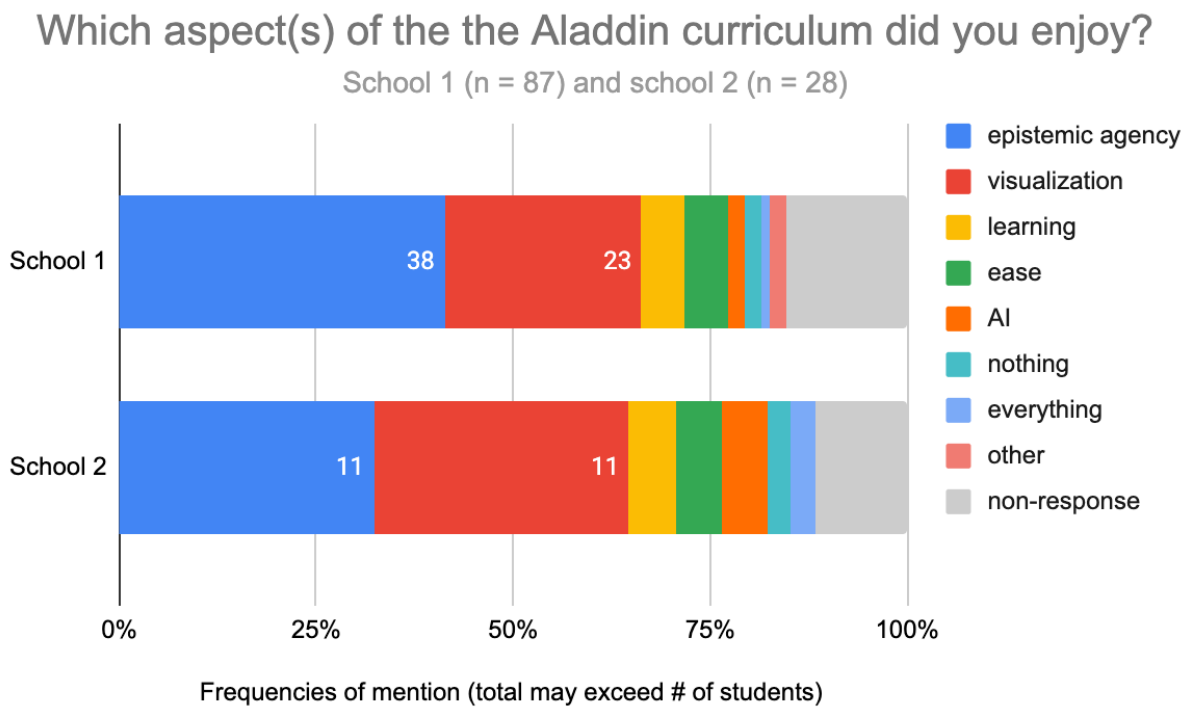
The teacher from school 2 found the AITA to be “real[ly] useful” in an informal conversation after the implementation, explaining that “in a lot of cases, the kids just need a suggestion.” He also suggested other types of feedback that the AITA could give, from more directive feedback (“you’ve tried to change the tilt angle 5 times, how’s it going for you so far? You haven’t touched the other variables”) to more facilitative feedback (“I see your number went down. Here’s some reasons it may have gone down instead of up like you expected”). On the other hand, the teacher observed off-task behaviors from his students, which the researcher also confirmed during observation, and reported having to maintain classroom discipline multiple times throughout the project.

Student feedback

In the free-response questions about the learning experience in the post-survey, students left both praises and criticisms of the Aladdin curriculum. On the positive side, 49 out of the 115 students surveyed mentioned the freedom and autonomy to design and experiment, which was related to their epistemic agency (Miller et al., 2018). For example, C_P6S2 stated: “I liked playing around with different factors and seeing how they affected the outcome, whether that was the revenue or the cost or the amount of kWh that were produced as I changed different things.” C_P6S28 was one of the 34 students that commented positively on the visual aspect of the Aladdin curriculum, stating that “once I saw [AITA’s] take on how to do it, it pointed me in the right direction to then make more profit and energy, which lead [sic] me to one of my final designs”. 4 students mentioned the AITA specifically, claiming that they enjoyed “the use of the AI to find the best design for the solar panel farm” (C_P4S16) because they “liked seeing the AI work and go through many [iterations] quickly” (C_P6S5).

On the other hand, 80 of the 115 students experienced some level of technological difficulty, reporting issues with freezing, lagging, or navigation. Students also provided

constructive criticism of the current implementation of AITAs. One student wished that “when the AI says the [design] parameters are off, it should explain how they are off. I should not have to figure it out” (C_P5S2). Another student wished to “really see the math behind the AI and have an explanation [of] why and how AI works” (C_P4S18). Finally, one student cautioned that “using an AI as a teaching method should be used sparingly and a human teacher should be used a majority of the time” (C_P4S33), with no specific explanation of their reasoning.



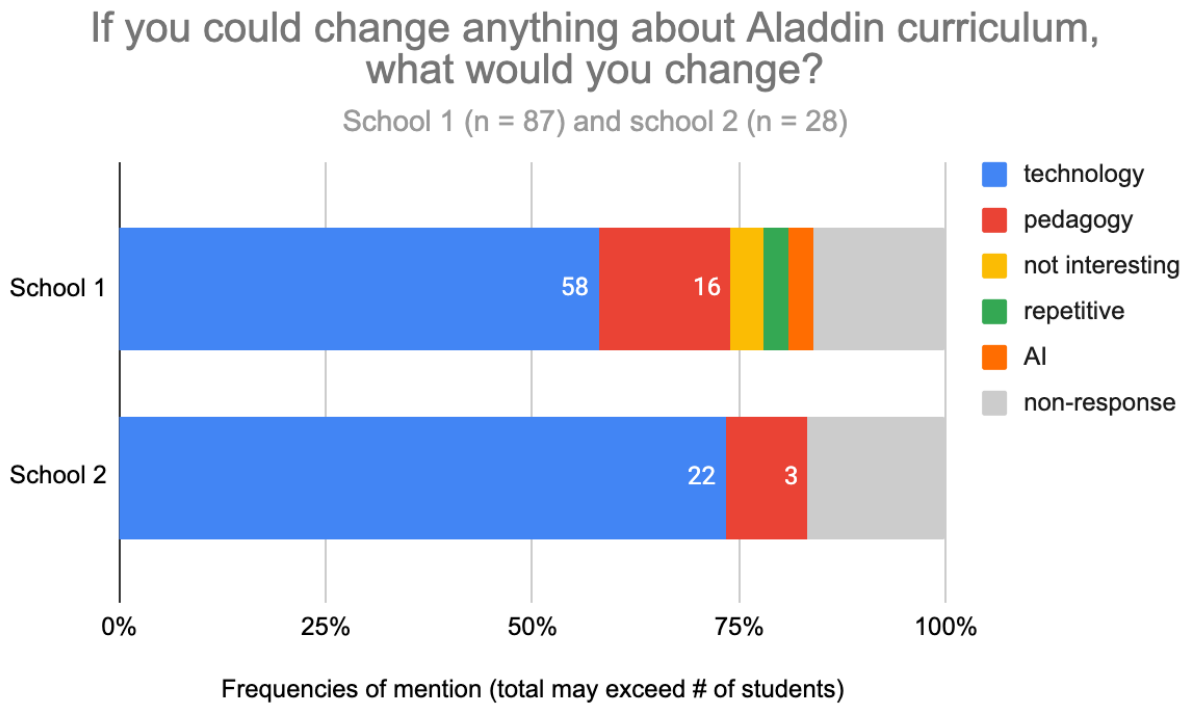


Figure 14. A breakdown of the student responses to the open-ended questions in the post-survey. (a) A question about what students liked about the curriculum. (b) A question about what students would change about the curriculum.

DISCUSSION

Science and engineering learning outcomes

Results from the multiple-choice questions in the pre- and post-tests suggested that students achieved some improved understanding of solar energy science concepts after finishing the solar farm design project. In two of the four questions (Q2 and Q4), the correct answers still didn't receive the majority vote in the post-test. Upon further examination of the question prompts and items, we determined that the questions themselves may be responsible for the results. The items in Q2 contained images that were prone to misinterpretation, and Q4

needed additional scaffolding. In addition, students may have applied the “goldilocks rule” of choosing the middle item to Q4 instead of applying the appropriate science concepts.

Results from the self-assessments also indicated that students achieved increased self-efficacy in engineering design after the project. When asked to identify important components of the engineering design process, students frequently mentioned more “hands-on” processes like “develop design solutions” and “construct a prototype”, a consistent observation across both tests that reflected the common impressions of engineering design among these students. They rarely mentioned processes like “identify a design need”, “select the best possible design”, and “communicate a design”, which was expected because those were not emphasized or practiced heavily in the project. Students in school 1 mentioned “research a design need” and “evaluate and test a design” more frequently than those in school 2 in the pre-test, but the differences were reduced in the post-test. This suggested that the solar farm design project may help students develop a more holistic impression of the engineering design process and reduce the achievement gap between different student populations. Finally, both schools saw significant increases in the popularity of “redesign” (including responses that mentioned “iterate”, “multiple designs”, or “trial and error”), which may in part be due to the interaction with the AITA and the extensive redesigning that many students went through.

Effects of the AITA

An analysis of the different types of AI feedback and student reactions showed that one of the most prominent effects of the AITA in Aladdin was either reinforcing or activating students’ existing knowledge, For the 10 of 16 students that correctly rejected AI feedback to decrease the tilt angle, their reflections showed that many already knew the optimal range of the tilt angle and therefore raised doubts when the AI feedback contradicted with the solar energy science knowledge they had already learned. As for the 3 of 16 students accepting AI feedback to increase the tilt angle, their reflections also showed that they were already aware of the optimal

range of tilt angle, but for whatever reason, they didn't optimize it to the fullest extent (35º) on their initial attempt. In this case, the AI feedback activated their existing knowledge and reminded them to further optimize it.

Another effect of the AITA was allowing students to create more divergent designs. The effect was most prominent in the case of the 3 students that went the extra mile to find a new RW-IRS pairing for their post-AI design. For the 8 students that simply adopted AI's exact RW-IRS pairing in their post-AI design, less could be said about whether the students learned anything about the coupled nature of RW and IRS, how they impacted design performance, and how divergent thinking could lead to better designs. However, it could be viewed as a case of collaborative intelligence between human and AI (Wilson & Daugherty, 2018), where each contributed what they knew about the design problem and took advice from each other.

On the other hand, C_P5S16 presented a rare case, where the AITA failed to improve the student's original design. Their reflection referred to the "initial testing" and "much more testing" that they had done to discover 2 of the local optima in the design space. Their activity log data also confirmed that they had already iterated 30 times before receiving AI feedback and another 17 times afterwards, which might explain why the student still managed to learn even from a failed attempt from AI: The student was very engaged and already explored the design space pretty thoroughly, giving them an edge in distilling helpful information from the raw feedback.

Finally, 2 students didn't further improve their designs after receiving AI feedback, likely due to not having iterated enough times. 2 other students reinforced their misconceptions about the tilt angle despite having improved their designs, likely due to changing multiple variables at a time instead of running controlled experiments. Combined with findings from the other cases, these observations suggested that the quality and extent of the effect of the AITA varied based on many factors such as the students' level of engagement, existing knowledge, engineering

design practices, etc., and that additional scaffolding may be necessary to support the learning of those students who were weaker in one or more of those aspects.

Pedagogical advantages of AITAs

A comparison of students' design performance before and after using the AITA, their design reflections, and their interviews provided both quantitative and qualitative data that indicated pedagogical advantages of using AITAs to provide feedback on student learning.

The use of AITAs is comparable to learning from peers (Wood & O'Malley, 1996). This was found to be a common theme in the interviews, with at least 5 out of 15 students identifying a need or preference for comparing their results with someone else's, be it another student or an AITA. In an in-person learning environment, students may spontaneously compare their design performance (yearly profit) with others. While this feedback isn't directly actionable without further comparison of design variables and process, it informs them of their relative standing in the group and what has been proven to be achievable in terms of design optimization. Similarly, the AITA provides a concrete design with quantitative performance data as a comparison, but it also exceeds typical peer feedback in two aspects: 1) While students sometimes only share their design performance but not their design variables, the AITA always provide both pieces of information, so that students can quantitatively compare both designs, identify specific changes that contribute to the improvement (if any), and act upon it. 2) While students rarely discuss their thought process or design rationale that lead to their designs, the AITA visualizes the design process in an animation, where students can see the evolution of one design iteration to the next. The abstract concept of divergence and convergence cycles can also be visualized: Students can see that the earlier design iterations look more different and the later iterations look more similar. This visual aspect facilitates learning from contrasting cases (Schwartz, 2016) and may be responsible for helping novice students adapt their design strategy from random trial and error to a more systematic approach.

The use of AITAs is also comparable to learning from experts. Assessing large numbers of engineering design solutions has always been a difficult task, especially for geoscience educators who may not have equal expertise in engineering design. The AITA provides an efficient approach: If the AI design performs much better than the student design, then it obviously means that the student design has lots of room for improvement. However, if the AI design doesn't show much improvement, then it means that the student design may already be close to the optimal solution. AI-generated feedback also exceeds expert feedback in two aspects: 1) While expert feedback is often based on design heuristics that still needs to be tested, AI-generated feedback is supported by quantitative evidence. 2) Student interview data suggests that the visualization of the computation process increases student confidence in the authority of the feedback, which may encourage adoption of such feedback.

Finally, there is evidence to support the hypothesis that the psychological effect of AITAs, especially the element of surprise, contributes to student learning. C_P4S30 documented how AI's surprising design prompted them to apply previously acquired knowledge about solar energy science to explain the surprise and identify ways to further improve the design. C_P4S14 explained how AI's surprising design triggered an important understanding of solar energy science: Seasonal changes of solar angles are an important factor in choosing an optimal tilt angle. These findings are consistent with the theory of cognitive development and constructivist learning (Lutz & Huitt, 2004): When presented with new knowledge that doesn't fit into any existing schema, the cognitive dissonance may trigger students to restructure their existing schema to *accommodate* the new knowledge.

Difference between schools

Student performance differed noticeably between the two schools. In multiple choice questions Q1 and Q3, school 1 outperformed school 2 in both the pre- and post-tests by a margin of 10%-20%. In Q2 and Q4, where school 2 outperformed school 1 in the pre-test,

school 2 actually performed worse in the post-test, whereas school 1 still improved. Given that both Q2 and Q4 had design flaws, the student performance at school 2 was more likely the result of guessing rather than reasoning.

Students from both schools reported similar levels of familiarity with engineering design in the pre-test. However, students from school 2 reported confidence levels that were consistently 5-15% lower than students from school 1 did across all engineering design processes. While the majority (54%) of students from school 1 reported the highest level of familiarity with engineering design in the post-test, only 21.4% of students did so in school 2. Similarly, even though both schools saw a roughly 20% overall increase in confidence levels in the post-test, students from school 1 still reported higher confidence levels for every design process.

The difference in engagement was also salient, considering that only 1 out of 31 students documented complete and valid AI data. Out of all students who submitted both the pre- and post-tests and had the opportunity to interact with AI in class, 12.5% (10 of 80) students from school 1 didn't document anything, while 64.2% (18 of 28) students from school 2 didn't document anything. These differences in self-efficacy and engagement may be inherent, given that the teacher from school 2 had commented on the behavioral issues in his classes, and there had been multiple accounts of observations where the teacher had to address such issues publicly in class. Nevertheless, improvements could be made to support the struggling students and address disengagement.

Room for improvement

Data from this preliminary study of AITAs in geoscience and engineering education generated important insight into how this pedagogical approach can be improved, specifically in the following aspects:

More informative feedback

Many students thought that the AITA could give “more specific” feedback. Recommendations collected from the students and teacher can be categorized as: (1) directive feedback, which includes pointing out what the student did wrong and telling the student explicitly what they need to do to improve the result; (2) facilitative feedback, which includes explaining the reason why the student didn’t perform as expected, asking the student to reflect on patterns in their process or behavior, or recommending that the student experiment with something new. It’s worth noting that students and the teacher preferred different types of feedback. Students generally wished for feedback that is immediately actionable, will quickly improve their design product, and require less cognitive effort, whereas the teacher recommended feedback that focuses more on scientific reasoning than immediate action, targets the design process or mindset more than the product, and requires more cognitive effort. Because students exhibit diverse needs that vary greatly depending on their levels of prior knowledge and current progress (Schwartz, 2016), it seems more desirable for AITAs to provide multiple types of feedback, rather than canonizing any one approach.

More psychological support

Student interview data also highlighted the importance of human interaction in the classroom. While the current AITA lacks the emotional attention of a teacher and the shared experience of a fellow student, and future research should explore how AI can support students in the psychological dimension in addition to the cognitive dimension, we would like to reiterate that the intention of this innovation is not to create any replacement for human interaction. Rather, it is to create an alternative to accommodate students’ diverse needs and a fallback in times of disruption.

More inclusive and equitable AI

There was some evidence that students who were more advanced in their studies or more engaged in the project benefited more readily from the current implementation of AITA in Aladdin, while students who were struggling or disengaged appeared to have gained less from

their interactions with the AITA. To make AI more inclusive and equitable, future research should focus on improving the usability of the software to reduce frustration and disengagement and providing both directive and facilitative AI feedback that address different student needs. To this end, large language models (LLMs) may be well suited for the task of providing natural language feedback in educational contexts.

Student perceptions of AITAs

Student perceptions of AITAs are mixed. In general, the AITA was regarded as being able to meet some students' needs, but not all. Students in school 1 rated the AITA more favorably than the students in school 2, which may be partially attributed to the fact that the students in school 2 exhibited more disengagement. These mixed responses agreed with a previous study that found variation in student perceptions of automated feedback (Calvo & Ellis, 2010). Student interview data further supports the positioning of AITAs not as a substitute for teacher/peer feedback, but as an additional source of feedback that can amplify the effect of human feedback (when available).

LIMITATIONS

There are several limitations of this evaluation. Firstly, though the pandemic was still ongoing at the time of the implementation, the two participating schools had both returned to in-person learning, with few social distancing guidelines in place. Therefore, the data collected does not fully represent the efficacy of the curriculum in an at-home online learning environment. Another limitation is that validated assessment items specific to solar energy engineering were not readily available, and improper assessment design may have influenced student responses to items like Q2 and Q4. Finally, it was difficult to assess and control for students' prior knowledge with solar energy or engineering design during the study, so some patterns in student responses to the AITA may be attributed to students' prior knowledge. For

example, students who had more familiarity with the subject matter may be more engaged in reflecting on the feedback they received and therefore exhibit more learning gains and a more positive attitude towards the AITA.

In terms of data analysis, the discussion of students' engineering design performance was based on a comparison of 3 snapshots within the entire design trajectory, namely the pre-AI, AI, and post-AI designs, instead of the evolution of all iterations. The potential correlations between factors such as engagement and performance were not discussed. In addition, the inter-rater reliability of coding and analysis of qualitative data was not measured due to there being only one researcher. These limitations could also provide directions for future research.

IMPLICATIONS

The viewing of the COVID-19 crisis as a catalyst for change is echoed by many in education (Zhao, 2020; Crick, 2021). The pandemic has exposed long-standing shortcomings in geoscience education that will likely outlast the COVID pandemic, such as inadequate integration with engineering design at the K-12 level, the lack of alternative practical opportunities for students with special needs, and the lack of alternative feedback mechanisms. However, with a wealth of experience accumulated over the past two years, geoscience educators are now in a much better position to proactively build a more equitable and resilient learning environment that cultivates student agency with regard to global challenges.

Part of the solution entails deeper integration of engineering design into geoscience education, which would afford students the opportunity to apply their science knowledge to solve pressing problems of today and tomorrow. ESS educators interested in using the solar farm design curriculum are welcome to explore the full problem space of solar energy engineering, a booming industry in demand of a greater workforce. For example, the profitability of the same solar farm design varies greatly in different geographic regions and depends on factors such as

the weather and the local electric rate. A profitable solar farm design in Northeastern US may not be profitable at all in Midwestern US due to less sunshine hours and lower electric rate, which can lead into rich discussions about the relationship among geoscience, engineering design, and public policy. In addition, Aladdin supports the use of custom ground images, which educators can use to overlay additional GIS data and discuss geological and environmental considerations during site assessment.

Similar to how online learning has transitioned from a novel concept to a common alternative during COVID, the other part of the solution is to introduce alternative forms of support into the learning environment, including selective use of AI. While the AITA in Aladdin remains available 24/7, educators interested in Aladdin and its accompanying curriculum materials should be mindful of providing multiple feedback mechanisms to accommodate diverse student needs, with AI feedback being an additional option that complements existing channels. Future research may focus on making AI-generated feedback more understandable, e.g. by incorporating strategies recommended in the engineering design coaching tool (Purzer et al., 2022). Another potential research direction is to extend the preliminary work on AI as instructional design agents with different personas (Schimpf et al., 2019) and create more humanized agents that are capable of assessing students' psychological state using their design activity data and providing socio-psychological interventions (Yeager & Walton, 2011), in addition to design feedback. As U.S. states begin to terminate the COVID-19 state of emergency and AI continues to gain traction in education, the findings about student interaction with an AITA in the time of COVID will only become more relevant for applying AI in ways that benefit not just some percentage of students, but all students.

ACKNOWLEDGEMENTS

The authors are indebted to the assistance of science teachers Derek Kendall and Russell Columbus, who allowed us to conduct research in their classrooms and implemented

the curriculum. The authors would also like to thank Dr. Alex Barco and Isaac Lyss-Loren for their contributions to data collection and analysis. This work was supported by the National Science Foundation (NSF) under grant numbers 2105695 and 2131097. Any opinions, findings, and conclusions or recommendations expressed in this paper, however, are those of the authors and do not necessarily reflect the views of the NSF.

DISCLOSURE STATEMENT

The authors report there are no competing interests to declare.

REFERENCE

Afzaal, M., Nouri, J., Zia, A., Papapetrou, P., Fors, U., Weegar, R., Wu, Y., & Li, X. (2021). Explainable AI for Data-Driven Feedback and Intelligent Action Recommendations to Support Students Self-Regulation. *Frontiers in Artificial Intelligence*. https://doi.org/10.3389/frai.2021.723447

Al-Shaery, A. M., Khozium, M. O., Farooqi, N. S., Alshehri, S. S., & Al-Kawa, M. A. M. B. (2022). Problem Solving in Crowd Management Using Heuristic Approach. *IEEE Access*, *10*, 25422–25434. https://doi.org/10.1109/ACCESS.2022.3156008

Bergey, P. (2020). Generative well pattern design—Principles, implementation, and test on OLYMPUS challenge field development problem. *Computational Geosciences*, *24*(6), 2079–2094. https://doi.org/10.1007/s10596-019-09912-w

Bethune, K. (2020). Changing Trends and Rethinking Geoscience Education in the Context of a Global Crisis. *Geoscience Canada*, *47*(4), 167–169. https://doi.org/10.12789/geocanj.2020.47.164

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.

Carberry, A. R., Lee, H.-S., & Ohland, M. W. (2010). Measuring Engineering Design Self-Efficacy. *Journal of Engineering Education*, *99*(1), 71–79. https://doi.org/10.1002/j.2168-9830.2010.tb01043.x

Carnegie Mellon University (n.d.). Prior Knowledge Self-Assessments - Eberly Center - Carnegie Mellon University. Retrieved March 3, 2023, from https://www.cmu.edu/teaching/assessment/priorknowledge/selfassessments.html

Calvo, R. A., & Ellis, R. A. (2010). Students' Conceptions of Tutor and Automated Feedback in Professional Writing. *Journal of Engineering Education*, *99*(4), 427–438. https://doi.org/10.1002/j.2168-9830.2010.tb01072.x

Crick, T. (2021). Covid-19 and digital education: A catalyst for change?. ITNOW, 63(1), 16-17.

Darvishi, A., Khosravi, H., Abdi, S., Sadiq, S., & Gašević, D. (2022). Incorporating Training, Self-monitoring and AI-Assistance to Improve Peer Feedback Quality. *Proceedings of the Ninth ACM Conference on Learning @ Scale*, 35–47. https://doi.org/10.1145/3491140.3528265

Goldin, I., Narciss, S., Foltz, P., & Bauer, M. (2017). New Directions in Formative Feedback in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education*, *27*(3), 385–392. https://doi.org/10.1007/s40593-016-0135-7

Gigerenzer, G. (2008). Why Heuristics Work. *Perspectives on Psychological Science*, *3*(1), 20–29. https://doi.org/10.1111/j.1745-6916.2008.00058.x

Haupt, G. (2015). Learning from experts: fostering extended thinking in the early phases of the design process. International Journal of Technology and Design Education, 25(4), 483-520.

Higgins, J. L. (1971). A New Look At Heuristic Teaching. *The Mathematics Teacher*, *64*(6), 487–495. https://doi.org/10.5951/MT.64.6.0487

Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial Intelligence for Assessment and Feedback to Enhance Student Success in Higher Education. *Mathematical Problems in Engineering*, *2022*, e5215722. https://doi.org/10.1155/2022/5215722

Hughes, B. B. (1974). Heuristic Teaching in Mathematics. *Educational Studies in Mathematics*, *5*(3), 291–299.

Lucas, J. F. (1974). The Teaching of Heuristic Problem-Solving Strategies in Elementary Calculus. *Journal for Research in Mathematics Education*, *5*(1), 36–46. https://doi.org/10.5951/jresematheduc.5.1.0036

Kim, J., Merrill, K., Xu, K., & Sellnow, D. D. (2020). My teacher is a machine: Understanding students' perceptions of AI teaching assistants in online education. International Journal of Human–Computer Interaction, 36(20), 1902-1911.

Kishore, P., & Kisiel, J. (2013). Exploring High School Students' Perceptions of Solar Energy and Solar Cells. *International Journal of Environmental and Science Education*, *8*(3), 521–534.

Lutz, S., & Huitt, W. (2004). Connecting cognitive development and constructivism: Implications from theory for instruction and assessment. Constructivism in the Human Sciences, 9(1), 67-90.

Massachusetts Department of Education (2001/2006). *Massachusetts science and technology/engineering curriculum framework.* Malden, MA: Massachusetts Department of Education.

Miller, E., Manz, E., Russ, R., Stroupe, D., & Berland, L. (2018). Addressing the epistemic elephant in the room: Epistemic agency and the next generation science standards. *Journal of Research in Science Teaching*, *55*(7), 1053-1075.

Mirchi, N., Bissonnette, V., Yilmaz, R., Ledwos, N., Winkler-Schwartz, A., & Maestro, R. F. D. (2020). The Virtual Operative Assistant: An explainable artificial intelligence tool for

simulation-based training in surgery and medicine. *PLOS ONE*, *15*(2), e0229596. https://doi.org/10.1371/journal.pone.0229596

National Research Council (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas.* Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

NGSS Lead States (2013). *Next Generation Science Standards: For States, By States.* Washington, DC: The National Academies Press.

Porter, B., & Grippa, F. (2020). A Platform for AI-Enabled Real-Time Feedback to Promote Digital Collaboration. *Sustainability*, *12*(24), Article 24. https://doi.org/10.3390/su122410243

Porter, S. G., Greene, K., & Esposito, M. C. K. (2021). Access and Inclusion of Students with Disabilities in Virtual Learning Environments: Implications for Post-Pandemic Teaching. *International Journal of Multicultural Education*, *23*(3), Article 3. https://doi.org/10.18251/ijme.v23i3.3011

Purzer, S., Schimpf, C., Quintana-Cifuentes, J., Sereiviene, E., Lingam, I., & Jiang, R. (2022). Refine By Design. The Science Teacher, 89(6).

Radford, J., Bosanquet, P., Webster, R., Blatchford, P., & Rubie-Davies, C. (2014). Fostering learner independence through heuristic scaffolding: A valuable role for teaching assistants. *International Journal of Educational Research*, *63*, 116–126. https://doi.org/10.1016/j.ijer.2013.02.010

Riggs, E. M. (2020). *Access and Success in Geoscience Education in the Age of COVID-19*. *2020*, ED027-05.

Schwartz, D. L., Tsang, J. M., & Blair, K. P. (2016). The ABCs of how we learn: 26 scientifically proven approaches, how they work, and when to use them. WW Norton & Company.

Shu, D., Cunningham, J., Stump, G., Miller, S. W., Yukish, M. A., Simpson, T. W., & Tucker, C. S. (2019). 3D Design Using Generative Adversarial Networks and Physics-Based Validation. *Journal of Mechanical Design*, *142*(7). https://doi.org/10.1115/1.4045419

Sung, S., Jiang, R., Huang, X., & Xie, C. (2022). Panels v. Trees: Broadening the Pathways of Engineering Education Through Integration With Social Studies. Science Scope, 45(6).

The United States Government. (2021, December 13). Fact sheet: President Biden signs executive order catalyzing america's clean energy economy through Federal Sustainability. The White House. Retrieved August 5, 2022, from https://www.whitehouse.gov/briefing-room/statements-releases/2021/12/08/fact-sheet-president-biden-signs-executive-order-catalyzing-americas-clean-energy-economy-through-federal-sustainability/

Thornton, C. P., Ruble, K., & Jacobson, L. A. (2022). Education for Children With Chronic Illness: Moving Forward in Online and Virtual Learning. *JAMA Pediatrics*, *176*(4), 341–342. https://doi.org/10.1001/jamapediatrics.2021.5643

U.S. Department of Energy. (2021, September). Solar Futures Study. Retrieved August 5, 2022, from https://www.energy.gov/eere/solar/solar-futures-study

U.S. Energy Information Administration. (2022). *Electric Power Monthly - U.S. Energy Information Administration (EIA)*. Electric Power Monthly. Retrieved August 5, 2022, from https://www.eia.gov/electricity/monthly/epm_table_grapher.php?t=table_es1a

van der Hoeven Kraft, K. J. (2017). Developing Student Interest: An Overview of the Research and Implications for Geoscience Education Research and Teaching Practice. *Journal of Geoscience Education*, *65*(4), 594–603. https://doi.org/10.5408/16-215.1

Vahdatikhaki, F., Salimzadeh, N., & Hammad, A. (2022). Optimization of PV modules layout on high-rise building skins using a BIM-based generative design approach. *Energy and Buildings*, *258*, 111787. https://doi.org/10.1016/j.enbuild.2021.111787

White, R., & Gunstone, R. (2014). Probing understanding. Routledge.

Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review*, *96*(4), 114-123.

Wood, D., & O'Malley, C. (1996). Collaborative learning between peers: An overview. Educational Psychology in Practice, 11(4), 4-9.

Xie, C., Schimpf, C., Chao, J., Nourian, S., & Massicotte, J. (2018). Learning and teaching engineering design through modeling and simulation on a CAD platform. *Computer Applications in Engineering Education*, *26*(4), 824–840. https://doi.org/10.1002/cae.21920

Xing, W., Li, C., Chen, G., Huang, X., Chao, J., Massicotte, J., & Xie, C. (2021). Automatic Assessment of Students' Engineering Design Performance Using a Bayesian Network Model. *Journal of Educational Computing Research*, *59*(2), 230–256. https://doi.org/10.1177/0735633120960422

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. Review of educational Research, 81(2), 267-301.

Yoo, S., Lee, S., Kim, S., Hwang, K. H., Park, J. H., & Kang, N. (2021). Integrating deep learning into CAD/CAE system: Generative design and evaluation of 3D conceptual wheel. *Structural and Multidisciplinary Optimization*, *64*(4), 2725–2747. https://doi.org/10.1007/s00158-021-02953-9

Zhao, Y. (2020). COVID-19 as a catalyst for educational change. Prospects, 49(1), 29-33.