Journal of Online Learning Research (2022) **8**(3), 289-313

Measuring Elementary Students' Behavioral Engagement in Web-based Science Inquiry Learning

JINGWEN HE
Beijing Normal University &
The Ohio State University
he.2093@osu.edu

BIHUI JIN
Beijing Normal University
bihuijin@mail.bnu.edu.cn

ZEXIN XU
The Ohio State University
jasonxu44@hotmail.com

DANHUI ZHANG
Beijing Normal University
danhuizhang@bnu.edu.cn

With the development of web-based science inquiry learning, behavioral engagement in such learning contexts received more and more attention. Combined with specific science inquiry stages: comparative experiment design, implementation with computer simulation, and reflection on results, the current study explored a series of features from log data to conceptualize students' behavioral engagement. The features were divided into three categories: general engagement features including time, game the system, submission frequency, and revisiting behavior; learning content related features including context consistency, comparative experimental design, and experiment design consistency; and instruction related features consisting of revision behavior and revision improvement. 220 sixth graders from four classes in China participated in the study. Correlation and regression analysis were used to analyze the relationship between engagement features and learning performance. The results showed that time spend in experimental implementation, gaming the system in experimental design and reflection, the number of trials and materials tried in experimental implementation, comparative experimental design and context consistency, and revision behavior in all stages was significantly correlated with the outcome variables. The regression analysis further indicated that revision behavior in experimental design was the most prominent predictor for the performance outcomes. The three kinds of features and the importance of feedback were further discussed.

INTRODUCTION

Web-based science inquiry is one of the most important activities to learn science (National Research Council, 2012). It encourages students to learn science in an authentic problem context. By generating research questions, making predictions, conducting experiments, collecting data, and drawing conclusions, students can solve the problem as scientists (Keselman, 2003). With the development of technology, computer supported science inquiry has been widely used (Pedaste et al., 2015). Computer simulation and visualization techniques allow students to conduct the virtual experiment and analyze data through graphs. In such way, students could connect scientific phenomena to graphical representations and promote their understanding of scientific concepts, graphical literacy, and inquiry skills (Donnelly-Hermosillo et al., 2020; Wen et al., 2020). In this study, Web-based Inquiry Science Environment (WISE), which has been widely used in K-12 science education worldwide (Raes & Schellens, 2016), was used to assist students' science learning.

Web-based science inquiry learning emphasizes students' self-regulation. Rather than passively receiving knowledge from a teacher's lecture, students actively engage in the learning activities to learn science and have more autonomy in their learning. Such learning context also requires students' higher self-regulated learning skills. Unlike a traditional class, there are more distractions in the web-based learning environment and students are more easily to be distracted (Bergdahl et al., 2020a). Previous studies have shown that students may exhibit disengagement behaviors such as gaming the system (Baker et al., 2004), carelessness (Pedro et al., 2014), and engaging in activities unrelated to learning tasks (Gobert et al., 2015). Therefore, it is necessary to monitor students' learning process and evaluate their engagement to promote web-based science learning.

To measure learner engagement, more and more researchers delve into using data generated by students during the learning process (Gobert et al., 2015). Log-data is a time-series record of students' activities in the online learning system (Henrie et al., 2018), including actions such as clicks, question responses, and page views with corresponding timestamps. The researchers extracted and modeled features of the log-data to measure students' learning engagement. Compared to self-report questionnaires and observations (Fredricks & McColskey, 2012), log-data objectively records student behavioral activities, providing valuable information for teachers and researchers to understand students' learning process.

The current study aims to extract features to measure behavioral engagement based on the web-based science inquiry learning process. Previous studies were concerned with students' engagement in a single course and used coarse-grained features. However, the current study focuses on the science inquiry context and generates fine-grained features, combining the engagement theory and the characteristics of scientific inquiry activities to assess students' behavioral engagement.

LITERATURE REVIEW

Science inquiry learning

The key elements in the science inquiry process

Inquiry-based learning refers to students learning science and constructing knowledge by proposing questions, making hypotheses, collecting data, and drawing conclusions in a complex problem context (Keselman, 2003). Pedaste et al. (2015) summarized five stages in science inquiry: orientation, conceptualization, investigation, conclusion, and discussion. Orientation refers to stimuli students' curiosity about the learning topic and emphasizes challenging tasks by problem statement; conceptualization means proposing questions or hypotheses based on theory, which includes two sub-stages of questioning and hypothesis generation. Inquiry refers to the process of exploring or experimenting, collecting and analyzing data according to an experimental design or exploration plan. It includes three stages exploration, experimentation, and data interpretation. Conclusion refers to drawing a conclusion from data and comparing the results with questions or hypotheses. The last stage, discussion, represents the presentation, reflection, and communication on the inquiry process or results. The current study focuses on the key elements in the inquiry process, which include experimental design, implementation, and reflection on the conclusion.

Web-based science inquiry learning

Though science inquiry learning has been recognized its significance worldwide, there are some challenges for students and teachers in practical teaching. Students may lack the strategy and knowledge to actively engage in science inquiry activities (Kruit et al., 2018). Students superficially follow the inquiry process while not genuinely involving deep thinking.

Teachers focus more on the learning content and knowledge rather than giving students opportunities to practice their ability to collect data, analyze data, and draw conclusions. What's more, teachers usually prefer more structural inquiry and give students less autonomy support (Lucero et al., 2013). Besides the challenge of teaching and learning, sometimes it is difficult to manipulate science experiments in class due to the limitation of the instrument, time, space, etc.

The development of computer-supported science inquiry meets the challenges in some ways (Pedaste et al., 2015). Science inquiry learning environment (ILEs) is a curricular system that supports science teaching, inquiry learning, embedded assessment, recording, and monitoring of students' learning process and outcome (Donnelly et al., 2014). ILEs provide students with meaningful and authentic inquiry problem scenarios that allow them to learn science in novel and interesting problem contexts. ILEs have powerful visualization techniques to support student learning. For example, students can use simulation models to conduct experiments and collect and analyze data through graphs and charts. ILEs support student collaboration and autonomous learning, especially in formulating hypotheses, collecting data, analyzing data, and drawing conclusions (Donnelly et al., 2014). Compared to traditional science class, web-based science inquiry class increases students' motivation to learn, makes science experiments more easily manipulated, and increase students' self-regulated learning.

Learning engagement

Engagement and learning performance

Learner engagement has been shown as an important indicator of performance and success (Lei et al., 2018), especially in the online learning context (Bergdahl et al., 2020b). Researchers usually divide engagement into three dimensions including behavioral engagement, cognitive engagement, and emotional engagement (Fredricks et al., 2004). Behavioral engagement emphasizes students' active participation in learning and academic tasks, including effort, persistence, focus, questioning, and active participation in discussions (Fredricks et al., 2004). Cognitive engagement emphasizes the mental effort that students put into the learning task. Students with high levels of cognitive engagement are not satisfied with a given learning task and tend to choose challenging tasks and use a variety of metacognitive strategies for learning (Fredricks et al., 2004). Emotional engagement refers to students' positive emotional responses to learning activities, teachers, and peers. The common affective responses in learning are concentration, confusion, and happiness (Ocumpaugh et al., 2015). Different researchers sometimes choose to focus on one or more aspects of learning engagement, depending on the needs of their study (Henrie et al., 2018). The current study focuses on students' behavioral engagement in science inquiry activities.

Previous studies have shown how engagement influence students' learning outcome (Lei et al., 2018). In the web-based learning environment, engagement is also positively related to students' performance (Bergdahl et al., 2020b) while disengagement behavior such as gaming the system is negatively related to performance (Baker et al., 2004). In science education, active engagement promotes students' understanding of complex science concepts (Pugh et al., 2010), academic performance, and inquiry skills (Lee et al., 2016; Wu & Wu, 2020). Wen et al. (2020) generated the frequency of hypothesis, frequency of experimental design, and implementation. However, they didn't find a significant correlation between these behavioral engagement features and science literacy. These results may indicate that the validity of these features needs further exploration.

Assessing engagement with log-data

Log data is a record of students' activities in the online learning system (Henrie et al., 2018). There are usually three types of studies that have been conducted to characterize learning engagement based on log data: (1) theory-driven approach to construct learning engagement measurement dimensions and indicators; (2) data-driven approach to construct learning engagement prediction models, and (3) direct approach to extract several behavioral indicators to characterize students' learning engagement.

Theory-driven studies emphasized the extraction of data indicators that build on existing educational theories (Fincham et al., 2019). They usually proposed a framework first and then conducted an exploratory factor analysis to explore the dimensional structure of the indicators and the validity of indicators under each input dimension. The following-up confirmative factor analysis was used to verify the structure of the indicators. However, such an approach failed to extract complex information about learning behavior from the rich log data, though it could provide the findings with good interpretability.

Data-driven studies usually adopted a supervised learning paradigm, using several features extracted from log data to predict the true value of a learning input label and construct a predictive model of the learning input. Researchers collect students' log data from the learning system, at the same time, they obtain the true value labels of learning inputs (i.e., the outcome variables of the prediction model) through video labeling, log replay coding, classroom observation, or self-report. After the feature extraction process, various supervised learning algorithms are used to train and validate the model, filter out valid predictors, and construct a learning input detector based on log data (D'Mello et al., 2017). Correlation and regression analyses are often used to validate the effectiveness of features (Henrie et al., 2018). However, such an approach focused more on the overall prediction accuracy, while failing to provide a good interpretation of each extracted feature.

In addition to the above two approaches, some researchers extracted features to characterize students' learning engagement using a variety of log-data. The underlying assumption is that the learning activities recorded in the log-file, such as time stamp and frequency, are positively related to engagement (Fincham et al., 2019). For example, in a gamified learning task, the researchers used the time spent on the mobile app and the frequency of task review to measure engagement. Moreover, different features, such as coin accumulation, coin trading, and time spent on treasure chests were extracted as behavioral indicators of disengagement (Syal & Nietfeld, 2020).

In the current study, we will directly extract the features from the log-file, but based on the theoretical framework of behavioral engagement in web-based science inquiry learning. Compared with other studies using the direct-extraction method (Fincham et al., 2019), by using the theoretical framework as our guidelines, we will provide a more comprehensive picture of engagement in the learning environment. Such a method overcome the limitations of using the pure theory-driven approach, which ignores the rich information provided by the log data. This method will overperform the traditional data-driven approach in that it will offer explicit explanations of the features.

Assessing engagement in science inquiry learning

Few studies have focused on assessing engagement in science inquiry learning. Wen et al. (2020) generated the frequency of hypothesis, experimental design, and implementation in CogSci for 8th graders. Gobert et al. (2015) used the following effective features to predict students' disengagement behavior in Inq-ITS: number of independent variable changes, the maximum time interval between an incomplete experiment and previous action, average pause time, the maximum time interval between actions, number of experiments not paused, the maximum time interval between running experiment and previous action.

There were also studies concerned with different stages of scientific inquiry. Peterson (2012) developed the Online Elements of Inquiry Checklist (OEIC) which divided science inquiry into eight phases: immersion, research questions, prediction, experimental design and procedure, observation, analysis and results, conclusion and interpretation, and future research and implications. The 40 elements were distributed across these eight stages, characterizing students' engagement at each stage. For example, in the conclusion and explanation stage, the elements were "Are the conclusions of the experiment connected to the data that was collected?", "Are the conclusions consistent with the data that was collected?," "Did the learners discuss the limitations of their research?" and "Did the learners justify their conclusions using data?", etc (Scogin & Stuessy, 2015, p.327). The percentage of completion of the elements in each stage was used as an indication of the level of engagement in that stage (Scogin & Stuessy, 2015).

Previous studies have paid less attention to the engagement in scientific inquiry learning (Sinatra et al., 2015). Science inquiry learning, as an important practical activity in science learning, have its unique characteristics. Many online learning engagement studies are based on coarse-grained learning behaviors (e.g., number of homework submissions, forum posts, time on the quiz), while the fine-grained behaviors of scientific inquiry activities (e.g., designing and implementing experiments) have not been considered.

Another limitation of previous studies lies in the sample. Most previous studies have focused on students' learning engagement in higher education learning systems (MOOC, learning management systems, distance learning systems, etc.), and have rarely paid attention to primary and secondary school students' engagement when using online learning systems. Therefore, the current study aims to measure elementary school students' engagement in web-based science inquiry learning.

METHODOLOGY

The current study aims to extract features to measure behavioral engagement based on the web-based science inquiry learning process and use these features to predict students' science learning performance. Two research questions were proposed:

RQ 1: What are the features to measure students' behavioral engagement in web-based science inquiry learning?

RQ 2: How do these features predict students' learning process performance and pre-and-post test scores?

Participants and process

The participants were 220 6th grade students from 4 classes in Inner Mongolia, China. Although none of the students had the experience of using the learning platform before, they could successfully get through with learning tasks after utility training. After excluding the missing value and the outliers, 211 students (106 males, 50.2%) were included in the final analysis.

The class was completed in three weeks. One week before the class, the 45-minute-pre-test was conducted. During the class, the researcher initially introduced the WISE platform to students in the first ten minutes so that students could know how they should conduct themselves on the WISE platform, and then the value and importance of teacher feedback were emphasized. After the training session, each student uses a computer to complete the learning project independently. During the three class periods, the researcher provides feedback to students twice after the first class and the second class. At the beginning of the second and third classes, the researcher asked students to read the feedback and revise their responses before continuing to learn. If students still have questions after reading the feedback, the researcher encouraged students to seek help, and she will provide face-to-face feedback to students. Within one week after the experiment, students spent around 45 minutes completing the post-test. No feedback was provided to the pre-and-post tests, nor the items covered in the learning material so that we can better observe the changes in students' knowledge integration ability. Permission was sought from the school principal and teachers before data collection and oral consent was also obtained from the participants.

WISE platform and learning material

Participants completed science inquiry learning through WISE (web-based inquiry science environment, http://wise.bnu.edu.cn) last for three learning hours. The unit used in this study was *Thermal Challenge*. The main question of this topic was "whether we could use the same material to design the thermos cup for hot or cold drinks." Students were asked to propose an experimental hypothesis, design experiments, carry out computer simulations, collect data, use the chart to represent data, and draw conclusions. NetLogo Models were embedded into the platform to help students understand how the drinks' temperatures change over time with different materials (see Appendix step 1.7). Students need to explore three experiments in total (Hot drink, Cold drink, the Mixed condition).

In this study, we focus on the hot drink experiment. Students were asked to finish step 1.6 to step 1.8 to go through the inquiry process. On step 1.6, students were required to fill out a table to design the controlled experiment. In the next step 1.7, students used computer simulation to experiment based on their previous design. On step 1.8, students were asked to answer an open-ended question to examine their understanding of the controlled experiments and graphing analysis. See the Appendix for detailed information about the three steps.

3.3 Log-data in the WISE platform

There were ten variables recorded in the log-data, including student ID, Component ID, Component Prompt, Server Timestamp, Log-type, Log content, Teacher Comment, Teacher score, and Teacher comment time stamp. Based on the learning content, multiple types of behaviors could be recorded, as shown in Table 1. All log content had a corresponding timestamp.

Table 1

Log-data recorded in the WISE platform

	Log type	Log Content	Meaning
1	nodeEntered	nodeEntered	Enter a certain page
2	nodeExited	nodeExited	Exit a certain page
3	goHomeSelected	goHomeSel	Exit the learning system
4	choiceSelected	e.g., A	Select an option in multiple choice
5	choiceSubmited	e.g., A	Submit an option in multiple choice
6	tableSubmited	e.g., aluminum 90 5	Submit a table for experimental design
7	openRespSubmited	e.g., My mom uses aluminum cups	Submit answer for an open response item
8	simulationRun	e.g., aluminum 90 5	Condition selected running computer simulation: aluminum-hot drink-low environment temperature
9	graphSubmited	e.g., wood 5 90	Submission of the graph result
10	graphDelete	e.g., wood 5 90	Delete a line in the graph

Learning performance

Pre-and-post tests

15 questions (7 items for pre-test, 8 items for post-test) were developed to examine students' scientific knowledge, ability to read and interpret charts and design controlled experiments. Most items were composed of two questions, one multiple-choice item followed by an open-ended response item, which asked students to further elaborate on their choice on the first question (Table 2). A rubric of open-ended response items was developed to assess students' capability in all four aspects of knowledge integration (Table 3), which is a process model of "generating ideas, adding ideas, using evidence to distinguish ideas, and building connections among ideas" (Gerard & Linn, 2016; Linn & Eylon, 2011).

Table 2
An example of the two-order questions

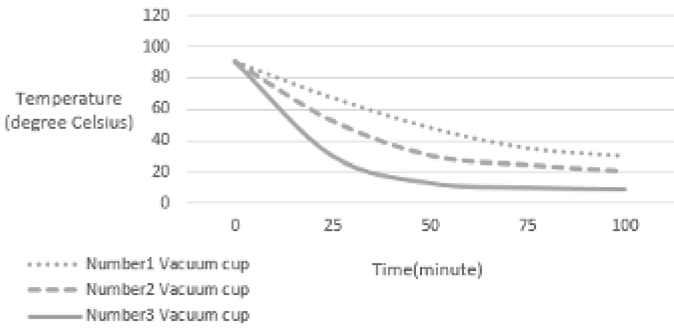
Question situation	The following picture shows the gradual cooling process of hot drinks in 3 different material cups over time. 
Question1	Which cup keep warm for the longest period? A. Number 1 B. Number 2 C. Number 3
Question2	Please use the evidence in the chart to explain your answer.

Table 3
An example of a five-point KI rubric on the item in Table 3

Score	Standard description	An example of students' answer
0	**No answer**	
1	**Off-task/incorrect:** The student's answer is not related to the question.	☐ I guess.
2	**Irrelative ideas:** Put forward relevant viewpoints but did not find the connection between viewpoints.	☐ The number 3 is the best, because it has the largest curvilinear curve.
3	**Partial Link:** Put forward relevant views, but it is not enough to solve the problem.	☐ The number 1 has the longest time to keep heat, because the temperature of the Number 1 thermos cup is always the highest.
4	**Full Link:** It can clearly explain a scientific and effective connection between the two viewpoints related to the given situation.	☐ The 3 thermos cups start at the same temperature, 100 minutes later, number 1 is about 30℃, number 2 is 20℃, and number 3 is about 10℃. So, number 1 is the best.
5	**Complex Link:** It can clearly explain two or more scientific and effective connections between multiple viewpoints of the given situation.	☐ The number 1 is the best. Because the temperature drop of Number 1 thermos cup is the gentlest. The initial temperature is the same, and after the same time, the temperature of number 1 thermos cup is the highest, so I choose this one.

Process performance

There were four embedded items on the platform: "Hot Drink Experimental Design", "Hot Drink Experimental Conclusion", "Cold Drink Experimental Design", and "Cold Drink Experimental Conclusion". The two experimental design questions asked students to fill out the form and were scored 0-1, and the two conclusion questions were open-ended with KI rubric from 0 to 5. We generated "experimental design ability" with "Hot Drink Experimental Design" and "Cold Drink Experimental Design" items. We also generated "graphing analysis ability" with "Hot Drink Experimental Conclusion" and "Cold Drink Experimental Conclusion" items.

RESULT

Features to measure students' engagement in science inquiry

The science inquiry process was divided into three major stages: controlled experiment design (step 1.6), implementation with computer simulation (step 1.7), and reflection on results (step 1.8), corresponding to the three key stages during the inquiry process. The following features were considered: time, game the system, submission frequency, revisit frequency, comparative experiment design (only in design and implementation stages), context consistency (only in design and implementation stages), experimental design consistency (only in implementation stage), revision behavior and revision improvement. These features could be further divided into three types: general engagement features, learning content related features and instruction related features. The explanation and coding for each feature was shown in Table 4.

Table 4

The explanation and coding for behavioral engagement features in science inquiry learning

	Sub-category	Explanation	Coding
General engagement features	time	Time spends on the page before teacher's feedback.	Continuous variable.
	game the system	Messy and meaningless answer before teacher's feedback. Students cannot go to the next step unless they finish all the questions in the current step. Some students may answer messy and meaningless answers such as "idk" or "ha-ha" to quickly go to the next step.	Binary variable.(1=game the system, 0=not game the system)
	submission frequency	Number of submissions before teacher's feedback. For implementation stage, number of computer simulation trails and number of materials tried before teacher's feedback are defined as submission frequency.	Continuous variable.
	revisiting behavior	Number of revisiting times before teacher's feedback. E.g. When designing experiment, student goes back to previous pages to acquire information.	Continuous variable.
Learning content related features	context consistency (Design and implementation)	Students need to set up drink temperature and environment temperature in experiment design. In hot drink experiment, the drink temperature should be higher than the environment temperature. Context consistency will be evaluated when student conduct comparative experiment. If student set up mixed condition, then context consistency is coded as missing. This feature is applicable in experimental design and implementation.	Binary variable. (1=consistent with the context, 0=not consistent with the context)
	comparative experimental design	Whether students control the drink temperature and environment temperature to be the same before teacher's feedback. Context consistency is not required. This feature is applicable in experimental design and implementation.	Binary variable. (1=control the condition to be the same, 0=mixed condition)
	experiment design consistency (implementation only)	Whether computer simulation results are consistent with previous table design. Students are supposed to conduct computer simulation experiment based on their previous design. This feature is applicable only in experimental implementation.	Binary variable. (1=consistent with previous table, 0=not consistent with previous table)
Instruction related features	revision behavior	Whether student revises the answer when the initial answer is incorrect.	Binary variable. (1=student revised the answer, 0=no revision)
	revision improvement	Students' improvement in their score. Only include students with revision behavior.	Continuous variable.

As students may change their behavior after receiving the teacher's feedback, only before feedback log-data was considered in terms of time, gaming the system, submission frequency, revisiting, context consistency, and comparative experiment design. Two feedback-related features, revision behavior and revision improvement were also included to examine how students use teacher's feedback.

Descriptive results of the features

The descriptive results of the features were shown in Table 5. There were eight features in the experimental design stage, ten features in the experimental implementation stage, and six features in the reflection stage.

Generally, students need around 4.9 minutes to fill out the design form, 3.16 minutes to run the computer simulation, and 3.48 minutes to finish the open-response question in the reflection stage. A few students performed gaming system behavior. More students answered meaningless responses in the reflection stage. Most students submit once in the experimental design and reflection stage. In the implementation stage, the number of trails and materials were used to represent submission frequency ranging from 0 to 13 and 0 to 6 respectively. Students tested around four trials and three materials on average. As for revisiting behavior, students showed more revisiting behavior in the implementation and reflection stages.

For Comparative experiment design and context consistency, less than half of the students were able to design or conduct a controlled experiment. Among these students, 71% and 92% identified the hot-drink context and set up the correct condition in the design and implementation stages, respectively. Interestingly, only 19% of the students followed their previous design in the implementation stage. 32 students designed the wrong table while drawing the correct graph in the later stage. This may indicate that students can self-regulated their learning with the assistance of computer simulation.

In terms of revision behavior, in the experimental design stage, there were 156 students who needed to revise their answers, among them 94 students revised their initial answer, and 75 students revised correctly. In the experimental implementation stage, there were 163 students who needed to revise the answer, while only 30 students showed revision behavior, and 14 students correctly revised the graph. In the reflection stage, 76 out of 148 students revised their answers, and 57 of them improved.

Table 5
Descriptive results of the engagement features

Science Inquiry Stage	Feature	N	Min	Max	Mean	SD
Controlled experiment design	1.Time (seconds)	199	2.00	793.00	295.16	161.24
	2.Game the system	210	0	1.00	0.04	0.19
	3.Submission Frequency	211	1.00	4.00	1.09	0.34
	4.Revisit Frequency	211	0	10.00	0.71	1.65
	5.Context consistency	78	0	1.00	0.71	0.46
	6.Comparative experiment design	196	0	1.00	0.40	0.49
	7.Revision Behavior	156	0	1.00	0.60	0.49
	8.Revision Improvement	94	0	1.00	0.80	0.40
Implementation with computer simulation	1.Time (seconds)	181	2.00	661.00	190.64	139.62
	2.Game the system	211	0	1.00	0.04	0.19
	3.Number of trails	211	0	13.00	4.11	2.59
	4.Number of materials	211	0	6.00	3.38	2.13
	5.Revisit Frequency	211	0	23.00	2.62	3.44
	6.Context consistency	76	0	1.00	0.92	0.27
	7.Experiment design consistency	206	0	1.00	0.19	0.40
	8.Comparative experiment design	166	0	1.00	0.44	0.50
	9.Revision Behavior	163	0	1.00	0.18	0.39
	10.Revision Improvement	30	0	1.00	0.47	0.51
Reflection on results	1.Time (seconds)	158	2.00	1366.00	209.36	224.24
	2.Game the system	210	0	1.00	0.10	0.30
	3.Submission Frequency	210	1.00	3.00	1.04	0.22
	4.Revisit Frequency	211	0	26.00	3.96	3.95
	5.Revision Behavior	148	0	1.00	0.51	0.50
	6.Revision Improvement	76	-1.00	5.00	1.50	1.28

How do features predict students' learning performance?

Table 6 to Table 8 showed the correlation results between features in three stages and students' performance, including both process performance and increased score in the pre-and-post test. In experimental design stage, gaming the system was negatively related to the integration score (r = -.198, p =.005), and revision behavior was positively correlated to all three levels of knowledge integration (from distinguish to integration, r = .168, *p* = .040; *r* = .211, *p* = .010; *r* =.178, *p* = .030). Gaming the system was also negatively related to the experimental design ability (*r* = -.207, *p* = .004) and graphing analysis ability (*r* = -.196, *p* = .006). Context consistency (*r* = .677, *p* < .001), comparative experimental design (*r* = .241, *p* = .001), revision behavior (*r* = .615, *p* < .001) and revision improvement (*r* = .744, *p* < .001) were positively related to the experimental design ability. Context consistency (*r* = .263, *p* = .022), comparative experimental design (*r* = .169, *p* = .021), and revision behavior (*r* = .315, *p* < .001) were also positively related to the graphing analysis ability.

Table 6

Correlation between features in experimental design stage and performance

	Time	Game the system	Submission frequency	Revisit	Context consistency	Comparative experimental design	Revision behavior	Revision improvement
Process performance								
Experimental Design Ability	.022	-.207**	-.063	-.034	.677**	.241**	.615**	.744**
Graphing Analysis Ability	.032	-.196**	-.069	-.021	.263**	.169*	.315**	.162
Pre- and Post- test improvement								
Distinguish	.010	-.061	-.010	-.013	-.070	-.021	.168*	.009
Integration	.047	-.198**	.025	.005	-.009	.006	.211**	.136
Complex Integration	-.024	-.092	-.017	.048	.051	-.043	.178*	.124

p<0.05 *; p<0.01 **; p<0.001 ***. Same below.

In experimental implementation stage, time was positively related to the complex integration skill (*r* = .181, *p* = .017). The number of trails was negatively correlated with distinguish skill (*r* = -.143, *p* = .042) and the context consistency was negatively correlated with complex integration (*r* = -.251, *p* = .032). Time (*r* = .221, *p* = .004; *r* = .233, *p* = .002), number of trails (*r* = .174, *p* = .016; *r* = .158, *p* = .027), number of materials (*r* = .328, p < .001; *r* = .216, *p* = .002), experiment design consistency (*r* = .318, *p* < .001; *r* = .235, *p* = .001), comparative experimental design (*r* = .272, *p* < .001; *r* = .211, *p* = .008) and revision improvement (*r* = .598, *p* < .001; *r* = .449, *p* = .015) were positively related to the experimental design ability and graphing analysis ability.

Table 7

Correlation between features in experimental implementation stage and performance

	Time	Game	Num of trails	Num of material	Revisit	Context	Experiment design	Compare experiment	Rev	Rev imp
Process performance										
Experimental Design Ability	.221**	-.127	.174*	.328**	-.070	.116.	.318**	.272**	.178*	.598**
Graphing Analysis Ability	.233**	-.142*	.158*	.216**	-.049	.070	.235**	.211**	.105	.449**
Pre- and Post- test improvement										
Distinguish	.036	.030	-.143*	-.084	-.024	-.195	-.035	.015	-.002	-.105
Integration	.027	-.090	-.001	.033	-.001	-.002	-.017	-.037	.038	.312
Complex Integration	.181*	-.069	-.075	.069	.056	-.251*	.001	-.084	-.022	.141

Game: game the system; Context: Context consistency; Experiment design: Experiment design consistency; Compare experiment: Comparative experimental design; Rev: Revision behavior; Rev imp: Revision improvement

In reflection stage, none of the features had significant correlation with pre- and post- test performance. Time (r = .223, p = .007), gaming the system (r = -.480, p < .001), revision behavior (r = .303, p < .001) and revision improvement (r = .682, p < .001) were found significantly related to the graphing analysis ability.

Table 8

Correlation between features in reflection stage and performance

	Time	Game the system	Submission frequency	Revisit	Revision behavior	Revision improvement
Process performance						
Experimental Design Ability	.014	-.270**	-.046	-.087	.056	.025
Graphing Analysis Ability	.223**	-.480**	-.104	-.047	.303**	.682**
Pre- and Post- test improvement						
Distinguish	-.011	-.073	-.057	-.027	-.007	-.120
Integration	-.019	-.062	-.125	-.002	-.102	.148
Complex Integration	-.148	-.118	-.079	.061	.028	-.002

Above all, among all the features, time spend in experimental implementation, gaming the system in experimental design and reflection, the number of trials and materials tried in experimental implementation, comparative experimental design and context consistency, and revision behavior in all stages had significant correlations with the outcome variables.

Regression analysis was used to examine how different features influence learning performance. As there were five outcome variables, five regression models were created. For each outcome variable, predictors with significant correlation were put into the model. For example, based on Table 5 to Table 7, pre-and-post test score in the distinguish dimension was significantly related to the revision behavior in the experimental design stage and the number of trails. Therefore, we put these two variables in the regression model.

As there were small variances in game the system, and a limited sample in the context consistency and revision improvement, they were not included in the regression analysis. As we can see from Table 9, **Revision behavior in experimental design** can significantly predict all three levels of pre-and-post test scores and two process performance. Gaming the system in experimental design can also predict students' integration scores.

Table 9

Regression analysis results

	Predictor	B	β	t	Sig.	R^2
Distinguish	Number of trials	-.031	-.139	-1.701	.091	**.047**
	Revision behavior in experimental design	**.234**	**.187**	**2.292**	**.023**	
Integration	Game the system in experimental design	-.566	-.189	-2.318	.022	**.076**
	Revision behavior in experimental design	**.213**	**.164**	**2.002**	**.047**	
Complex Integration	**Revision behavior in experimental design**	**.269**	**.178**	**2.191**	**.030**	.032
Experimental Design Ability	Comparative experimental design in experimental design	-.021	-.018	-.163	.871	**.279**
	Comparative experimental design in experimental implementation	.080	.064	.565	.574	
	Experiment design consistency	.117	.089	.726	.471	
	Revision behavior in experimental design	**.428**	**.481**	**4.29**	**<.001**	
	Revision behavior in experimental implementation	.114	.095	.848	.400	
	Time spent on experimental implementation	.001	.202	1.385	.171	
	Number of trails	.000	-.002	-.012	.990	
	Number of materials	-.034	-.138	-.869	.388	

Table 9, *Continued*

	Predictor	B	β	t	Sig.	R^2
Graphing Analysis Ability	Comparative experimental design in experimental design	.428	.160	1.168	.249	**.230**
	Comparative experimental design in experimental implementation	.484	.213	1.362	.180	
	Experiment design consistency	.533	.140	1.011	.317	
	Revision behavior in experimental design	**.788**	**.355**	**2.096**	**.042**	
	Revision behavior in reflection	.317	.160	1.028	.309	
	Time spent on reflection	.000	-.018	-.131	.896	
	Number of trials	.101	.225	1.123	.267	
	Number of materials	-.126	-.221	-.901	.372	

DISCUSSION

The current study explored elementary students' behavioral engagement in the web-based science inquiry environment. Responding to our first research question, three types of features were extracted from log-data: **the general engagement features** including time, gaming the system, submission frequency (number of trails and number of materials), revisiting behavior; **the learning content related features** including context consistency, comparative experimental design, and experiment design consistency; and the **instruction related features** including revision behavior and revision improvement.

Regarding the second research question, correlation and regression analysis were used to validate the effectiveness of the extracted features. Correlation analysis showed that time spend in experimental implementation, gaming the system in experimental design and reflection, the number of trials and materials tried in experimental implementation, comparative experimental design and context consistency, and revision behavior in all stages was significantly correlated with the outcome variables. The regression analysis further indicated that revision behavior in experimental design was the most prominent predictor for the performance outcomes.

The first interesting finding worth discussion was why revision behavior in experimental design could significantly predict all the outcome variables. This might be related to the importance of feedback and the essential role of the experimental design stage. Teacher's feedback is an important instructional strategy (Azevedo & Hadwin, 2005) that helps students scaffold inquiry, promotes active engagement of prior knowledge, monitors their learning process and goals, and uses more effective learning strategies (Azevedo et al., 2005; Khishfe & Abd-El-Khalick, 2002). It plays an essential role in helping students monitor their self-regulated learning progress, directing them to focus on their learning goals and ultimately improving their learning outcomes (Ozan & Kıncal, 2018; Xiao & Yang, 2019). In the web-based learning environment, students need to use self-regulated learning strategies to handle the challenges in the learning tasks. Teacher feedback tells students the gap between the current performance and ultimate goal (Hattie & Timperley, 2007), and also provides them with opportunities to reflect on their performance. When students engage in the feedback, they could realize their weaknesses and direction for the next step. However, the current study also indicated that only around 60% of the students revised their answers. Therefore, a more detailed and well-developed feedback design should be considered in future studies.

The feedback during the experimental design stage was particularly essential in science inquiry learning (Pedaste et al., 2015; Schwichow et al., 2022). Experimental design guides the following implementation and reflection (Pedaste et al., 2015). When students design the experiments, they propose the specific questions and goals, make hypotheses, design inquiry plans, and choose the data collection method (Hodson, 2014). One of the aims of the current learning project was to develop students' VOTAT strategy (vary-one-thing-at-a-time), which means they could only change the materials each time and control other influencing factors, the temperature of drink and environment, to be the same. Consistent with the previous studies, though many students could understand the concept of VOTAT strategy, it becomes challenging when they need to design experiments by themselves (Schwichow et al., 2022). In sum, if students could receive teacher's feedback and make revisions, it is very likely that they could correct their own initial misunderstanding and redesign the experiment.

The second interesting finding was the correlation between the general engagement features during the experimental implementation stage and process performance. The significant general engagement features included time, gaming the system, number of trials, number of materials, and revision behavior. Students who spend more time interacting with computer simulations, showing less gaming system behavior, and trying more materials and conditions tend to have better performance. The unique learning tool in the experimental implementation stage was the computer simulation, which has powerful visualization techniques that allow students to collect and analyze data through graphs and charts (Cui et al., 2022). Students set up the experiment conditions first, click the "run" button, and they could wait and see the change of temperature in the graph (See Appendix step 1.7). The more experiments students run, the more time they will spend. Similarly, the more numbers of materials they tried, the more time and effort they will need to interpret the results. Therefore, the log-data during the experimental implementation stage indeed reflect students' effort in science inquiry learning.

The learning content related features, including comparative experimental design, context consistency, and experimental design consistency, were also significantly correlated with the process performance. These features were closely related to the objective of the current learning project. The comparative experimental design was related to the VOTAT strategy. The context consistency was related to recognition of the problem context. For example, in a hot drink experiment, the drink temperature should be higher than the environment temperature. The experimental design consistency was related to following the guidance of the instruction material, which refers to whether computer simulation results are consistent with the previous experiment design. We found that some students who were able to use VOTAT strategy to design the experiment failed to recognize the problem context. Therefore, whether students could show coherent thinking about comparative experimental design and context consistency may reflect the difference in their science inquiry skills and knowledge integration skills.

CONCLUSION AND IMPLICATION

The current study explored elementary students' behavioral engagement in experimental design, implementation, and reflection stages in the web-based science inquiry environment. The general engagement features, the learning content related features, and the instruction related features were extracted from the log-data. Correlation and regression analysis were used to validate the features. Revision behavior in experimental design was found to be the most prominent predictor for the performance outcomes. Time spends in experimental implementation, gaming the system in experimental design and reflection, number of trials and materials tried in experimental implementation, comparative experimental design and context consistency, and revision behavior in all stages was significantly correlated with the outcome variables.

This study complements the field of behavioral engagement in K-12 STEM education. As most of the previous studies focused on students in college, this study informs instructional design and teacher preparation programs in the elementary and secondary school levels. Also, the COVID epidemic has provided an urgency to reevaluate, revise, and reinvent traditional K-12 instructional delivery for more effective and viable alternatives. This study of virtual learning for elementary students is extremely important for establishing research-based practices and policies for online delivery of instruction at the K-12 level. Another implication is that the current study goes beyond the general features such as time and frequency, and targets to specific inquiry learning behavior. As engagement is a situated performance which relies on the learning context, our study enriches the concept of engagement and provides guidance for web-based inquiry science environment.

As for the future direction, there are several aspects that could be stressed. First, the current study only includes three science inquiry stages in the analysis. Although these three stages are the key elements in science inquiry, other elements such as communication and discussion could also be considered. Second, more advanced statistical methods or learning analytics techniques such as cluster analysis and sequential pattern analysis can be used to compare students in different groups. Third, the combination of different features and the relationship between features can be examined. For example, the combination of comparative experimental design and context consistency may indicate the nuance of students' knowledge integration skills. It is also interesting to investigate the causes of the 'game the system' feature, exploring the strategies for mitigating the occurrence of negative learning behaviors. Last but not least, multiple types of data such as think aloud, screen recording, and facial expressions can be collected to enrich the data resource.

CONFLICT OF INTEREST

The authors confirm that there are no known conflicts of interest with respect to this research, authorship, and publication.

APPENDIX

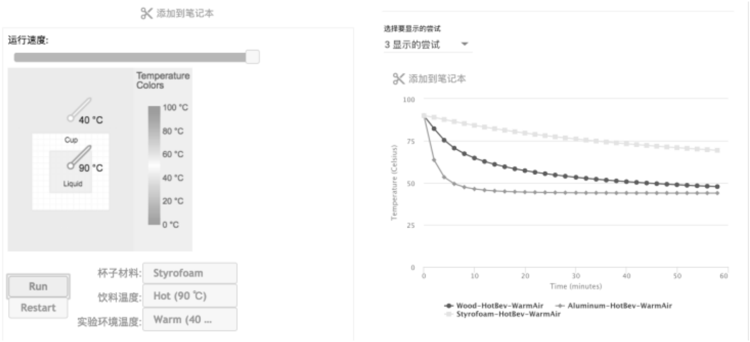
Step 1.6 to step 1.8 on the WISE platform

Step 1.6: Please use the condition below to fill out the form to prepare for your online experiment.

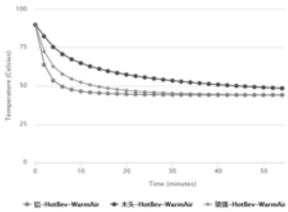
Materials: aluminum, wood, foam plastics, clay, glass, plastic.
The temperature of drink: hot, warm, cold.
The temperature of the environment: hot, warm, cold.

Trial	Materials	Temperature of drink	Temperature of environment
1			
2			
3			
4			
5			
6			

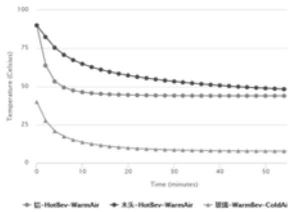
Step 1.7: Please choose the conditions based on your design. Run the model and compare the lines in the graph.



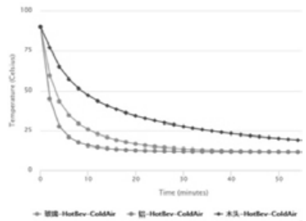
Step 1.8: After the computer simulation, we can see the lines in the graph. How can we know which material can best preserve the heat? Below shows the results from three students. Their research question is “which material can best preserve the heat? Aluminum, wood, or glass?” Combined with your design, please comment on their experiment.



Student A



Student B



Student C

References

Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition-Implications for the design of computer-based scaffolds. *Instructional Science, 33*(5), 367-379.

Azevedo, R., Cromley, J. G., Winters, F. I., Moos, D. C., & Greene, J. A. (2005). Adaptive human scaffolding facilitates adolescents' self-regulated learning with hypermedia. *Instructional Science, 33*(5), 381-412.

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students "game the system." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 383–390).

Bergdahl, N., Nouri, J., & Fors, U. (2020a). Disengagement, engagement and digital skills in technology-enhanced learning. *Education and Information Technologies, 25*(2), 957–983.

Bergdahl, N., Nouri, J., Fors, U., & Knutsson, O. (2020b). Engagement, disengagement and performance when learning with technologies in upper secondary school. *Computers in Education, 149*, 103783.

Cui, Y., Zhao, G., & Zhang, D. (2022). Improving students' inquiry learning in web-based environments by providing structure: Does the teacher matter or platform matter? *British Journal of Educational Technology, 53*(4), 1049-1068.

D'Mello, S., Dieterle, E., & Duckworth, A. (2017). Advanced, analytic, automated (AAA) measurement of engagement during learning. *Educational Psychologist, 52*(2), 104–123.

Donnelly, D. F., Linn, M. C., & Ludvigsen, S. (2014). Impacts and characteristics of computer-based science inquiry learning environments for precollege students. *Review of Educational Research, 84*(4), 572–608.

Donnelly-Hermosillo, D. F., Gerard, L. F., & Linn, M. C. (2020). Impact of graph technologies in K-12 science and mathematics education. *Computers in Education, 146*, 103748.

Fincham, E., Whitelock-Wainwright, A., Kovanović, V., Joksimović, S., Staalduinen, J.-P. van, & Gašević, D. (2019). Counting clicks is not enough: Validating a theorized model of engagement in learning analytics. In P*roceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 501–510).

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: a comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement (*pp. 763–782). Springer.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. R*eview of Educational Research, 74*(1), 59–109.

Gerard, L. F., & Linn, M. C. (2016). Using automated scores of student essays to support teacher guidance in classroom inquiry. Journal of Science Teacher Education, 27(1), 111-129.

Gobert, J. D., Baker, R. S., & Wixon, M. B. (2015). Operationalizing and detecting disengagement within online science microworlds. *Educational Psychologist, 50*(1), 43–57.

Hattie, J., & Timperley, H. (2007). The power of feedback. R*eview of educational research, 77*(1), 81-112.

Henrie, C. R., Bodily, R., Larsen, R., & Graham, C. R. (2018). Exploring the potential of LMS log data as a proxy measure of student engagement. *Journal of Computing in Higher Education, 30*(2), 344–362.
Hodson, D. (2014). Learning science, learning about science, doing science: Different goals demand different learning methods. *International Journal of Science Education, 36*(15), 2534-2553.
Keselman, A. (2003). Supporting inquiry learning by promoting normative understanding of multivariable causality. *Journal of Research in Science Teaching, 40*(9), 898–921.
Khishfe, R., & Abd-El-Khalick, F. (2002). Influence of explicit and reflective versus implicit inquiry-oriented instruction on sixth graders' views of nature of science. *Journal of Research in Science Teaching, 39*(7), 551–578.
Kruit, P. M., Oostdam, R. J., Berg, E. van den, & Schuitema, J. A. (2018). Effects of explicit instruction on the acquisition of students' science inquiry skills in grades 5 and 6 of primary education. *International Journal of Science Education, 40*(4), 421–441.
Lee, C. S., Hayes, K. N., Seitz, J., DiStefano, R., & O'Connor, D. (2016). Understanding motivational structures that differentially predict engagement and achievement in middle school science. *International Journal of Science Education, 38*(2), 192–215.
Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis. *Social Behavior and Personality, 46*(3), 517–528.
Linn, M. C., & Eylon, B. S. (2011). *Science learning and instruction: Taking advantage of technology to promote knowledge integration.* Routledge.
Lucero, M., Valcke, M., & Schellens, T. (2013). Teachers' beliefs and self-reported use of inquiry in science education in public primary schools. *International Journal of Science Education, 35*(8), 1407–1423.
National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* National Academies Press.
Ocumpaugh, J., Baker, R. S., & Rodrigo, M. M. T. (2015). Baker Rodrigo Ocumpaugh *Monitoring Protocol (BROMP) 2.0 technical and training manual.* NY, NY: Teachers College, Columbia U. Manila, Philippines: Ateneo Laboratory for the Learning Sciences.
Ozan, C., & Kıncal, R. Y. (2018). The effects of formative assessment on academic achievement, attitudes toward the lesson, and self-regulation skills. *Educational Sciences: Theory & Practice, 18*(1), 85-118.
Pedaste, M., Mäeots, M., Siiman, L. A., Jong, T. de, Riesen, S. A. N. van, Kamp, E. T., … Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review, 14*(1), 47–61.
Pedro, M. O. C. Z. S., Baker, R. S. J. de, & Rodrigo, M. M. T. (2014). Carelessness and affect in an intelligent tutoring system for mathematics. *International Journal of Artificial Intelligence in Education, 24*(2), 189–210.
Peterson, C. A. (2012). *Mentored engagement of secondary science students, plant scientists, and teachers in an inquiry-based online learning environment* (dissertation). Texas A&M University, College Station, TX.
Pugh, K. J., Linnenbrink-Garcia, L., Koskey, K. L. K., Stewart, V. C., & Manzey, C. (2010). Motivation, learning, and transformative experience: A study of deep engagement in science. *Science Education, 9*4(1), 1–28.
Raes, A., & Schellens, T. (2016). The effects of teacher-led class interventions during technology-enhanced science inquiry on students' knowledge integration and basic need satisfaction. *Computers in Education, 92*, 125–141.

Schwichow, M., Brandenburger, M., & Wilbers, J. (2022). Analysis of experimental design errors in elementary school: how do students identify, interpret, and justify controlled and confounded experiments? *International Journal of Science Education, 44*(1), 91-114.

Scogin, S. C., & Stuessy, C. L. (2015). Encouraging greater student inquiry engagement in science through motivational support by online scientist-mentors. *Science Education, 99*(2), 312–349.

Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational psychologist, 50*(1), 1-13.

Syal, S., & Nietfeld, J. L. (2020). The impact of trace data and motivational self-reports in a game-based learning environment. Computers in Education, 157, 103978.

Wen, C. T., Liu, C. C., Chang, H. Y., Chang, C. J., Chang, M. H., Chiang, S. H. F., … Hwang, F. K. (2020). Students' guided inquiry with simulation and its relation to school science achievement and scientific literacy. *Computers in Education, 149*, 103830.

Wu, P. H., & Wu, H. K. (2020). Constructing a model of engagement in scientific inquiry: Investigating relationships between inquiry-related curiosity, dimensions of engagement, and inquiry abilities. *Instructional Science, 48*(1), 79–113.

Xiao, Y., & Yang, M. (2019). Formative assessment and self-regulated learning: How formative assessment supports students' self-regulation in English language learning. System: An International *Journal of Educational Technology and Applied Linguistics, 81*, 39-49.