JOURNAL
OF BALTIC
SCIENCE
EDUCATION

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

UTILIZATION OF PROCESS DATA IN CHINA: EXPLORING STUDENTS' PROBLEM-SOLVING STRATEGIES IN COMPUTER-BASED SCIENCE ASSESSMENT FEATURING INTERACTIVE TASKS

Pingping Zhao,
Chun-Yen Chang,
Yueyang Shao,
Zhi Liu,
Hao Zhou,
Jian Liu

Abstract. *Students' problem-solving strategies and the differences among strategy groups were explored by analyzing the process data collected during student interactions with computer-based science items. Data were gathered from 1516 eleventh-grade students from 4 schools in China. Analyses of the sequences of students' response actions revealed that the students were divided into four strategy groups when designing experiments to solve scientific problems: the scientific and rigorous strategy (18.5%), scientific and less rigorous strategy (25.4%), incomplete strategy (31.5%), and chaotic strategy (24.6%). The heatmaps of response actions for each strategy and the frequencies of the most representative response sequences were further explored to understand the students' detailed trajectories. The results showed that successful problem solvers were generally inclined to explore all possibilities of experimental combinations and design experiments scientifically and rigorously based on the relevant scientific principles. Moreover, the timestamps of response actions were explored to show that the students who adopted the scientific and rigorous strategy spent more time seeking solutions, suggesting that students may need sufficient time to solve complex and authentic scientific problems. The findings enrich the literature on using process data to address theoretical issues in educational assessment and provide students with individualized instructional needs for teachers to improve students' scientific problem-solving competency.*
Keywords: *process data, scientific problem-solving, computer-based assessment, China*

Pingping Zhao
Beijing Normal University at Zhuhai, China
Chun-Yen Chang
National Taiwan Normal University, Taiwan
State University of Malang, Indonesia
Yueyang Shao, Zhi Liu, Hao Zhou, Jian Liu
Beijing Normal University, China

Introduction

The rapid growth of science and technology has led to fundamental and momentous changes in the societies of the 21st century. These changes are reflected in the types and complexities of problems encountered in everyday life. To successfully address these problems, problem-solving competency that focuses on in-depth exploration, critical thinking, and analytical reasoning in the process of understanding and applying knowledge is necessary and crucial for individuals (Lavoie, 1993; NRC, 1996). Problem-solving competency is an essential foundation for students to actively participate in social activities and engage in meaningful learning in the 21st century (NRC, 2010; OECD, 2013). The ability to solve real-world problems is a crucial component of scientific literacy and a significant goal of science education (NRC, 1996; NRC, 2012). Scientific problem-solving competency can help students construct a systematic knowledge structure and fully apply knowledge to cope with the challenges caused by the rapid development of science and technology (Friege & Lind, 2006; NRC, 1996).

To better cultivate students' ability to solve scientific problems, the processes of solving problems and students' struggles to seek solutions need to be explored. Recently, computer-based assessments that support human-computer interaction have emerged in various problem-solving domains (Jiang et al., 2021; Vendlinski & Stevens, 2002). Computer-based assessments record detailed interactions between the problem solver and the task environment and thereby capture salient solution processes (Chung & Baker, 2003; Provasnik, 2021). The rich process data generated from the interactive items reflect the processes of answering the items and provide tremendous opportunities to reveal the cognitive and behavioral processes of the problem solvers. A growing number of studies have focused on process data to reflect students' cognitive processes during problem solving. For

https://doi.org/10.33225/jbse/23.22.929

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

example, some attempts have been made to explore the patterns of log files obtained from the National Assessment of Educational Progress (NAEP) and the Programme for International Student Assessment (PISA) (Bergner & von Davier, 2019; Greiff et al., 2015).

However, it is still challenging to derive specific implications from the behavioral patterns found in log files because the cognitive process during problem solving is a black box and it is difficult to make sense of the hundreds of pieces of information students produce when solving problems in computer-based assessments (Greiff et al., 2015). Therefore, further exploration of process data is necessary to obtain more substantial findings, which can inform teachers and policymakers about students' cognitive processes during problem solving. Meanwhile, limited research has been performed to emphasize the implementation of independently developed interactive assessments in China and the exploration of Chinese students' cognitive processes of scientific problem solving. Hence, process data were utilized to explore cognitive strategies that eleventh-grade students adopt when solving interactive scientific tasks developed in China.

Literature Review

Scientific Problem-solving Competency

Problem-solving competency determines how successful an individual will be at finding solutions to challenges in life (Lazakidou & Retalis, 2010; OECD, 2013). Mayer proposed a widely accepted viewpoint that problem solving is the cognitive process of transforming a given situation into a target situation when there is no obvious solution (Mayer, 1992). The PISA 2012 problem-solving framework also stated that problem-solving competency is the ability of an individual to transform a given problem situation into a target situation by understanding the problem and solving it when the solution is not obvious and the individual's emotional willingness to participate in the problem solving (OECD, 2013). Hence, researchers emphasize that problem solvers explore the paths of solving problems from the initial state to the final goal state. Problem solving is a complicated process filled with continuous exploration from an initial state to a target state, which may involve multiple cognitive and emotional variables (OECD, 2013; Reid & Yang, 2002).

Consistent with the aforementioned viewpoints, researchers have stated that problem solving in the field of science education is a process of scientific inquiry to some extent (Gayford, 1989). Students need to have some scientific knowledge and process skills to solve scientific problems successfully (Chang et al., 2007). The process of solving scientific problems involves the generation and explanation of ideas (Chang & Weng, 2002). Students need to think divergently to seek solutions to solve problems and need to explain and evaluate various solutions to choose the final one. Although researchers have considered scientific problem-solving competency from various perspectives, they are generally aware that scientific problem solving is a process of inquiry that requires a series of scientific process skills. Therefore, scientific problem-solving competency in this study involves using a series of scientific process skills combined with scientific knowledge to seek solutions and evaluate them to solve complex problems relevant to science and technology in daily life.

Assessment Framework for Scientific Problem-solving Competency

Researchers have stated that there is domain-general problem solving that pays attention to problem situations across domains and domain-specific problem solving that focuses on specific educational areas, such as mathematics and science (Greiff & Neubert, 2014; Sternberg, 1995; Sugrue, 1995). Although the scientific problem-solving competency in this study is domain-specific problem solving, the studies of domain-general problem solving have also been instructive for us to consider our assessment framework because the two lines of research show considerable overlap. One of the most well-known assessments in the area of problem solving is the PISA 2012, which utilized a framework for domain-general problem solving (Greiff et al., 2014). This framework underlines a computer-based test design in which the students need to deal with daily real-life problems such as configuring an MP3 player. Students must explore and understand the problem situation (exploring and understanding), represent the problem by constructing representations, formulate hypotheses by identifying the relevant factors in the problem (representing and formulating), devise and carry out a plan to reach the goal state (planning and executing), monitor their progress and reflect on solutions from various perspectives (monitoring and reflecting) (OECD, 2013). Therefore, problem-solving competency involves formulating a hypothesis, devising and carrying

https://doi.org/10.33225/jbse/23.22.929

out a plan to solve the problem, and monitoring and critically evaluating problem-solving solutions, which are attributes of scientific inquiry.

Some studies put problem solving in a special, domain-specific domain based on its context. Polya originally proposed a four-stage problem-solving model and revised it in mathematics education, including understanding the problem, devising a plan, trying and carrying out the plan, and monitoring and reflecting on the solution (Polya, 1957). The four-stage problem-solving model is consistent with the problem-solving framework in the PISA 2012 to some extent. After Polya, some studies developed a problem-solving framework and assessment in the field of science. Some researchers have proposed that problem solving in science education has four overlapping and interactive aspects: problem posing, problem approaches, problem solutions, and communication (McIntosh, 1995). Problem posing includes extracting a science problem from a realistic situation, making or revising a hypothesis, and planning experiments. Problem approaches involve controlling variables in an investigation, collecting appropriate data, and revising approaches when warranted by new evidence. Problem solutions include developing various solutions to the same problem, comparing the results, and making conclusions based on data. Communication focuses on communicating procedures, interpretations, and thinking pathways. Hence, the scientific problem-solving process has obvious scientific inquiry attributes and involves various scientific process skills (Eysenck & Keane, 2000; Gayford, 1989; Ross & Maynes, 1983). Moreover, scientific discovery as dual search (SDDS), an important model utilized to explain the process of scientific problem solving (Schauble, 2003), perceives scientific problem solving as a coordinated search in two problem spaces, the hypothesis space and experiment space (Klahr & Dunbar, 1988). In the hypothesis space, the initial state is domain knowledge, and the target state is hypotheses. Search in the experiment space is directed toward experiments that discriminate between rival hypotheses and yield interpretable outcomes (Eysenck & Keane, 2000). It is obvious that scientific problem solving is a process of scientific inquiry to some extent (Gayford, 1989). Based on what scientists do when designing experiments and solving scientific problems, problem solving involves various skills, including formulating hypotheses, designing experiments, recording information, and judging the data collected (Ross & Maynes, 1983). Researchers' emphasis on experimental problem solving further indicates that the design of inquiry experiments is a crucial component of scientific problem-solving competency.

Reasoning is another component of scientific problem-solving competency. Educational researchers in the earth and space sciences claim that problem-solving skills include domain-specific knowledge, reasoning skills, and attitudes (Chang et al., 2007). Researchers in chemistry state that reasoning ability is highly related to students' ability to solve stoichiometric problems successfully (Robinson & Niaz, 1991). The PISA2012 problem-solving assessment framework also clearly puts forward the importance of reasoning, arguing that problem-solving processes rely on one or more reasoning skills (OECD, 2013). Additionally, explanations for problem solutions necessary for evaluating proposals are perceived as another component of scientific problem-solving competency. Studies suggest that problem solvers tend to articulate the steps to a problem solution and why they choose this solution (Chang & Weng, 2002; Chi et al., 1994). Successful problem solvers produce self-explanatory and principle-based explanations to effectively relate the solution to the principles or knowledge in science (Camacho & Good, 1989).

Three components, including the design of scientific inquiry, scientific reasoning, and scientific explanation, are perceived as crucial dimensions in the assessment framework for scientific problem solving in this study.

Use of Process Data to Reflect Students' Cognitive Processes in Computer-based Assessments

It is essential to understand the students' cognitive processes in solving scientific problems and to identify students who are struggling with these processes for further instruction (Bergner & von Davier, 2019; Greiff et al., 2015). With the conversion of science assessments from paper-and-pencil to computerized formats, process data derived from interactive item types emerged (Jiang et al., 2021; Provasnik, 2021). Log files are data sources for process data that reflect the cognitive and metacognitive processes involved in completing test tasks (Provasnik, 2021). Process data provide fine-grained information about how students plan, select, and execute various problem-solving strategies to find a solution. Therefore, the crucial role of process data is to delineate how students solve problems, not just whether they do. Especially when process data are combined with theoretical frameworks and response data, abundant opportunities are afforded to study problem solvers' paths to a solution (Provasnik, 2021).

Process data mainly include detailed records of student interactions with the computer system, such as action sequences when solving simulation-based science tasks, answer change behaviors, and the timestamps of these actions (Bergner & von Davier, 2019). For computer-based science assessment focusing on interactive tasks, an

https://doi.org/10.33225/jbse/23.22.929

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

important piece of process data is the sequences of problem solvers' response actions. These sequences provide insights into the strategies problem solvers frequently apply to solve scientific problems. For example, researchers exploited log files containing the process data of problem-solving behavior obtained from the 2012 cycle of the PISA (Greiff et al., 2015). Log-file analyses were conducted to explore whether the vary-one-thing-at-a-time (VOTAT) strategy that students employed while solving the climate control item was related to their performance and then to identify several groups of students according to their exploration strategies. The classification of response strategies based on the analyses of problem solvers' action sequences is a good way to identify classes of problem solvers and their specific needs (Arslan et al., 2020; Jiang et al., 2021).

Another crucial piece of process data that researchers pay attention to is the timestamps of response actions. Extensive studies have explored total response time on test questions as an indication of motivation (Lee & Jia, 2014; van der Linden, 2008). The timestamp data obtained from NAEP have been used to show the speed at which students compose written responses to test prompts and reflect various patterns and amounts of writing (Bergner & von Davier, 2019; Provasnik, 2021). Moreover, some research indicated that a long response execution time might suggest that the problem solvers are stumped by a problem; a short execution time might represent their high proficiency or rapid guessing (Guo et al., 2016; Lee & Jia, 2014). Other researchers claim that successful problem solvers devote considerable time to understanding the problem situation and seeking solutions (Gong et al., 2020). When constructing solutions to problems, experts generally devote sufficient time to effectively develop a strategy for creating a solution that is different from that of novices, who often approach problems haphazardly (Yerushalmi & Eylon, 2015). The aforementioned conclusions drawn on the analyses of response time are different and even opposite. This study can provide further information about the patterns of students' cognitive processes when solving scientific problems.

Problem Statement

As mentioned earlier, due to the complexity of the scientific problem-solving process, there is much room for improvement in analyzing students' cognitive strategies when solving scientific problems. By categorizing response strategies, categories of problem solvers and their specific needs can be identified, and this provides valuable reference and evidence for educational intervention to improve problem solvers' scientific problem-solving competency (Arslan et al., 2020; Jiang et al., 2021). However, there has been little research on Chinese students' classification of cognitive strategies for solving scientific problems from the perspective of process data. Therefore, the purpose of the present study was to explore the classification of students' experimental design strategies when solving scientific problems and the differences across strategy classes by analyzing process data. Accordingly, this study was conducted by analyzing Chinese students' sequences of actions and response times with a focus on three research questions:

RQ1: What is the classification of experimental design strategies that eleventh-grade Chinese students apply when solving scientific problems?
RQ2: What is the detailed trajectory of each class of experimental design strategies?
RQ3: Is there a significant difference in response time across strategy classes?

Research Methodology

Research Design

The methodology employed in this study was that of a quantitative analysis. The computer-based science assessment was administered to student participants in July 2021, and quantitative data were obtained. Process data including the logs of students' actions and the corresponding timestamps were analyzed. Descriptive statistics and Analysis of Variance (ANOVA) revealed the patterns of students' strategies when solving scientific problems. First, descriptive statistics of students' actions were performed to present a classification of experimental design strategies. Second, descriptive statistics were performed on students' response sequences to reveal the detailed trajectories of each category of experimental design strategies. Finally, to explore the differences in response time across strategy classes, descriptive statistics and ANOVA tests on response time were conducted.

https://doi.org/10.33225/jbse/23.22.929

Participants

In this study, a convenience sample of four high schools from a city (abbreviated as "M") located in southwest China was recruited. The sample schools, including two urban and two suburban schools located in four regions in M city, provide ordinary educational resources to students. All four schools were in the middle rank (top 50.0%) of the education quality in M city. The students who attended the schools could be described as belonging to urban and suburban areas, with the school serving low, middle, and upper-income families. Students in grade eleven from the four sample schools participated in the test at the end of the spring term of the 2020-2021 academic year, which constituted a total sample of 1516 students (57.0% female, 43.0% male). These students were between the ages of 16 and 17. The process data of students' response actions to the inquiry task in subtask 1 of the Handy Freezer (abbreviated as "HF," see Figure 1) task needs to be collected to analyze their problem-solving strategies. This inquiry task requires students to continually click buttons and record the experimental data. Students' response actions were recorded by the computer systems every 100 milliseconds as process data to show the real-time change in students' responses. Eventually, the process data of 1516 students' action sequences and response times were recorded for the analyses.

Instrument

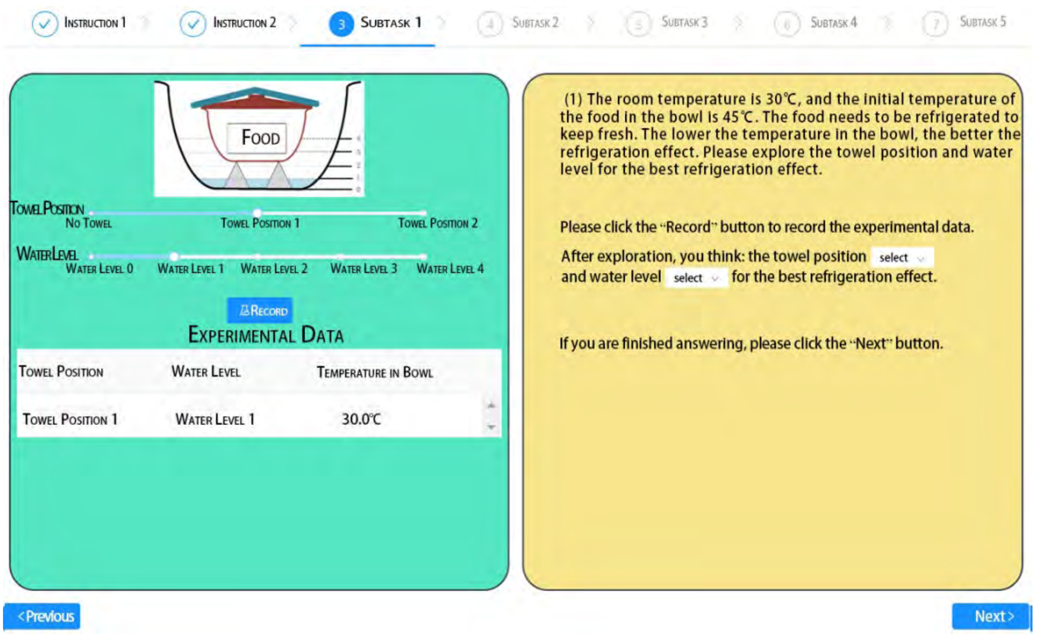
A computer-based scientific problem-solving assessment featuring interactive tasks was developed. Completion of tasks required only basic computer skills such as clicking on virtual buttons and sliders. The assessment instrument was developed based on the aforementioned framework of scientific problem-solving competency containing three elements (design of scientific inquiry, scientific reasoning, and scientific explanation). The instrument included three tasks consisting of several subtasks and covered the subjects of Physics, Chemistry and Biology. The Physics task consisted of five subtasks; the Chemistry task was made up of six subtasks; and the Biology task consisted of six subtasks. The programming tool JavaScript was utilized to form tasks, and the tasks were embedded in the test system OpenCT (https://open-ct.com). OpenCT is an interactive evaluation system based on Metaverse and educational big data for assessing higher-order thinking skills. The tasks were developed by a research team of three researchers (the first, third and fourth authors) and two technicians. Two of the three researchers were skilled in designing tasks associated with scientific literacy assessment. The other researcher was proficient in assessment theory. The two technicians were adept at designing interfaces for interactive tasks. One task was original and belonged to biology. The other task, belonging to physics, was adapted from an item developed by a research group from Guangxi Normal University. This item was included in a large-scale paper-and-pencil assessment to assess students' scientific literacy in mainland China. Another task, belonging to chemistry, was retrieved from the work of a research team at the National Taiwan Normal University. The instrument in this study was revised according to the comments from three rounds of student interviews and two rounds of expert reviews. Each round of student interviews included eighteen eleventh-grade students; the first round of expert reviews contained three professors who were good at assessing scientific literacy; and the second round of expert interviews included four skilled professors in designing science tasks. The Cronbach's alpha for this instrument is .863. The comparative fit index (CFI), Tucker-Lewis index (TLI), and root mean square error of approximation (RMSEA) are key indicators of model fitness. The values for validity (CFI= .82, TLI= .80, RMSEA= .08) of this instrument show that the model is a moderate fit (Karadakal et al., 2015).

Students had 60 minutes to complete the test. This study analyzed the process data collected from student interactions with one inquiry task in subtask 1 of the HF task (see Figure 1). The Chinese version of this inquiry task was translated into English for reading convenience. The inquiry task involved assessing the ability to design scientific inquiry included in the framework of scientific problem-solving competency and evaluating students' knowledge on the scientific content area of evaporation in physics. On this task, students were instructed to click the buttons representing the "Water Level" and "Towel Position" to explore the condition for the best refrigeration effect. To solve this task, they needed to (1) click the "Water Level" button, (2) click the "Towel Position" button, (3) click the "Record" button to record the experimental data for the selected combination of "Water Level" and "Towel Position," and (4) continuously explore in this way and draw a preliminary conclusion about the condition for the best refrigeration effect according to the experimental data. Detailed logs of the actions and the corresponding timestamps were recorded in-process data and used for analysis.

https://doi.org/10.33225/jbse/23.22.929

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Figure 1
Subtask 1 of a Task Named Handy Freezer



Measures

Response actions. The sequences of actions on the inquiry task in subtask 1 were used to classify the problem-solving strategies according to the principles of designing experiments. The overall numbers of each experimental combination within each strategy group and the frequencies of the most representative response sequences were analyzed to explore the concrete trajectory of each strategy group. In this study, clicking the button representing the "Towel Position" and "Water Level" and recording the experimental data was conceived as a set of response actions recorded as log data. There were three towel positions and five water levels (see Table 1). Specifically, "Towel Position 1" meant the towel covered the lid of the bowl, and "Towel Position 2" meant the towel was long enough to touch the bottom of the bowl. There were fifteen combinations of variables in total (see Table 2). The combination of "Towel Position 2" and "Water Level 2" (code [2, 2]) was the condition for the best refrigeration effect. We classified the problem solvers' response actions based on the principles of designing controlled experiments to better understand students' problem-solving strategies.

Table 1
Variables and Corresponding Codes

Variables	Specific Circumstance	Corresponding Codes
Towel Position	No Towel	0
	Towel Position 1	1
	Towel Position 2	2
Water Level	Water Level 0	0
	Water Level 1	1
	Water Level 2	2
	Water Level 3	3
	Water Level 4	4

https://doi.org/10.33225/jbse/23.22.929

Table 2
Experimental Combinations and Corresponding Codes

Experimental Combinations	Corresponding Codes
No Towel and Water Level 0	[0, 0]
No Towel and Water Level 1	[0, 1]
No Towel and Water Level 2	[0, 2]
No Towel and Water Level 3	[0, 3]
No Towel and Water Level 4	[0, 4]
Towel Position 1 and Water Level 0	[1, 0]
Towel Position 1 and Water Level 1	[1, 1]
Towel Position 1 and Water Level 2	[1, 2]
Towel Position 1 and Water Level 3	[1, 3]
Towel Position 1 and Water Level 4	[1, 4]
Towel Position 2 and Water Level 0	[2, 0]
Towel Position 2 and Water Level 1	[2, 1]
Towel Position 2 and Water Level 2	[2, 2]
Towel Position 2 and Water Level 3	[2, 3]
Towel Position 2 and Water Level 4	[2, 4]

Score. The students' total scores on all three interactive tasks in this assessment and scores on the HF item were indicators of problem-solving ability. To explore the effectiveness of the classification of strategies, total scores on the assessment and scores on HF items were recorded and used for further analysis. Regarding scores on HF items consisting of five subtasks, the full credits for these five subtasks were 1, 8, 8, 1, and 6, respectively. Taking subtask 1 as an example, when one student selected "Towel Position 2" and "Water Level 2" (the correct answer), full credit 1 was given to this student. Additionally, the maximum possible score on the HF item was 24 based on the sum of full credits for five subtasks, and the maximum possible total test score was 76 according to the sum of full credits for all three interactive tasks in this assessment.

Response Time. The time-based measures mainly include the execution time and total response time. Execution time elapsed between the first and the last actions on the inquiry task in subtask 1 of the HF item, reflecting the time needed to perform a mental plan and execute response actions. Total response time is the total time spent on the HF item.

Data Analysis

Quantitative analyses, including descriptive statistics and ANOVA tests, were carried out to address the three research questions. Descriptive statistics were conducted to compare the aforementioned process-related measures across strategy groups. It should be noted that the data of the numbers of clicks for each experimental combination for each strategy group were generated into a two-dimensional array format via Python, and then the Matplotlib Library was applied to visualize data to heatmaps. To validate the classification of strategy groups, ANOVA tests on total test scores and HF item scores were conducted via the IBM Statistics SPSS 24.0 software. To explore the potential differences in the problem-solving processes among students using different strategies, the ANOVA tests were performed on the categorical variables (e.g., total response time, execution time).

https://doi.org/10.33225/jbse/23.22.929

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Research Results

Classification of Experimental Design Strategies

When designing experiments to solve scientific problems, controlling the variables in the experiment is crucial to keep the other variable(s) constant and change only one variable under investigation across conditions. Additionally, even if a student can control variables scientifically, it does not mean that the student can design the entire experiment rigorously. According to the principles of controlling variables and designing experiments, students' responses were classified into four strategies. This study has two variables: "Towel Position" and "Water Level." Table 3 presents the descriptions of four strategies and their corresponding response actions.

Table 3
Descriptions of Four Strategies

Strategies	Descriptions of Strategies	Corresponding Response Actions	Representative Codes
Scientific and rigorous (SR)	Able to scientifically control variables and design rigorous experiments	(1) Be able to keep the variable "Towel Position" constant and collect five sets of "Water Level" data, and at least two sets of "Towel Position" can be controlled	[1, 0], [1, 1], [1, 2], [1, 3], [1, 4]; [2, 0], [2, 1], [2, 2], [2, 3], [2, 4]
		(2) Be able to keep the variable "Water Level" constant and collect three sets of "Towel Position" data, and at least two sets of "Water Level" can be controlled	[0, 0], [1, 0], [2, 0]; [0, 1], [1, 1], [2, 1]
Scientific and less rigorous (SLR)	Able to scientifically control variables and design less rigorous experiments	(1) Be able to keep the variable "Towel Position" constant and collect five sets of "Water Level" data, but only one "Towel Position" can be controlled	[2, 0], [2, 1], [2, 2], [2, 3], [2, 4]
		(2) Be able to keep the variable "Water Level" constant and collect three sets of "Towel Position" data, but only one "Water Level" can be controlled	[0, 0], [1, 0], [2, 0]
Incomplete (IN)	Able to control variables but design incomplete and nonrigorous experiments	(1) Be able to keep the variable "Towel Position" constant, but collect incomplete data (only four or three sets of "Water Level" data)	[1, 1], [1, 2], [1, 3], [1, 4]
		(2) Be able to keep the variable "Water Level" constant, but collect incomplete data (only two sets of "Towel Position" data)	[0, 2], [1, 2]
Chaotic (CH)	Unable to control variables and design chaotic experiments	Unable to control variables, and collect data chaotically and irregularly	[1, 2], [0, 0], [2, 1]

To validate the classification, students' total scores received on the assessment and on the HF item for each strategy group were compared. The results for total scores in Table 4 show that the students who adopted the SR strategy received the highest scores. In contrast, the students who used the CH strategy received the lowest scores, suggesting that the total scores declined with the emergence of the nonrigorousness of the experimental design when solving scientific problems. Additionally, for the SR strategy, the maximum value of total scores was 49.00, while the maximum value for the CH strategy was 34.00. Similar trends were found for scores on HF items. The students who applied the SR strategy got the highest scores on the HF item, while the ones who adopted the CH strategy received the lowest scores. Furthermore, the maximum value of scores on the HF item for the SR strategy was 20.00, while the CH strategy had the highest value of 16.00.

https://doi.org/10.33225/jbse/23.22.929

Table 4
Descriptive Statistics of Total Scores and Scores on HF Items across Strategy Groups

Measure	Scientific and Rigorous		Scientific and Less Rigorous		Incomplete		Chaotic	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Total Scores	23.79	10.24	19.43	10.07	17.17	9.62	10.21	7.03
Scores on HF Item	7.23	4.12	5.86	4.02	5.03	4.13	2.52	2.89

The differences in total scores and scores on HF items across strategy groups were further explored. Results of a one-way ANOVA on total scores showed that the means of the four strategies differed (F (3, 1512) = 124.573, $p < .05$). Games-Howell was utilized to run post hoc tests because the assumption of homogeneity of variance was violated according to the results of the Levene statistic in this study. Pairwise comparisons of the means using Games-Howell tests indicated four significant comparisons: the mean score for the SR strategy (M = 23.79, SD = 10.24) was significantly ($p < .05$) larger than that for the SLR strategy (M = 19.43, SD = 10.07); the mean score for the SLR strategy (M = 19.43, SD = 10.07) was significantly ($p < .05$) larger than that for the IN strategy (M = 17.17, SD = 9.62); the mean score for the IN strategy (M = 17.17, SD = 9.62) was significantly ($p < .05$) larger than that for the CH strategy (M = 10.21, SD = 7.03); and the mean score for the CH strategy (M = 10.21, SD = 7.03) was significantly ($p < .05$) smaller than that for the SR strategy (M = 23.79, SD = 10.24). Similarly, results for scores on HF item suggest that the means of the four strategies were unequal according to a one-way ANOVA, F (3, 1512) = 90.213, $p < .05$. Pairwise comparisons of the means using Games-Howell tests indicated four significant comparisons: the mean score for the SR strategy (M = 7.23, SD = 4.12) was significantly ($p < .05$) larger than that for the SLR strategy (M = 5.86, SD = 4.02); the mean score for the SLR strategy (M = 5.86, SD = 4.02) was significantly ($p < .05$) larger than that for the IN strategy (M = 5.03, SD = 4.13); the mean score for the IN strategy (M = 5.03, SD = 4.13) was significantly ($p < .05$) larger than that for the CH strategy (M =2.52, SD = 2.89); and the mean score for the CH strategy (M = 2.52, SD = 2.89) was significantly ($p < .05$) smaller than that for the SR strategy (M = 7.23, SD = 4.12). These results suggest that the total scores and scores on HF items were significantly different across strategy groups, implying that the classification of strategies can be confirmed experimentally. The strategies adopted when exploring and solving problems could predict problem-solving performance.

Regarding the percentage of each strategy group in this study, the results showed that 18.5% of the problem solvers adopted the SR strategy. A total of 25.4% of the students applied the SLR strategy, and 31.5% adopted the IN strategy. A total of 24.6% of the students designed experiments chaotically and irregularly. Since the SR strategy is a more efficient way to solve scientific problems than other strategies, approximately 81.5% of eleventh-grade problem solvers need to improve their ability to design scientific and rigorous experiments to successfully solve scientific problems.

Detailed Trajectory of Each Experimental Design Strategy

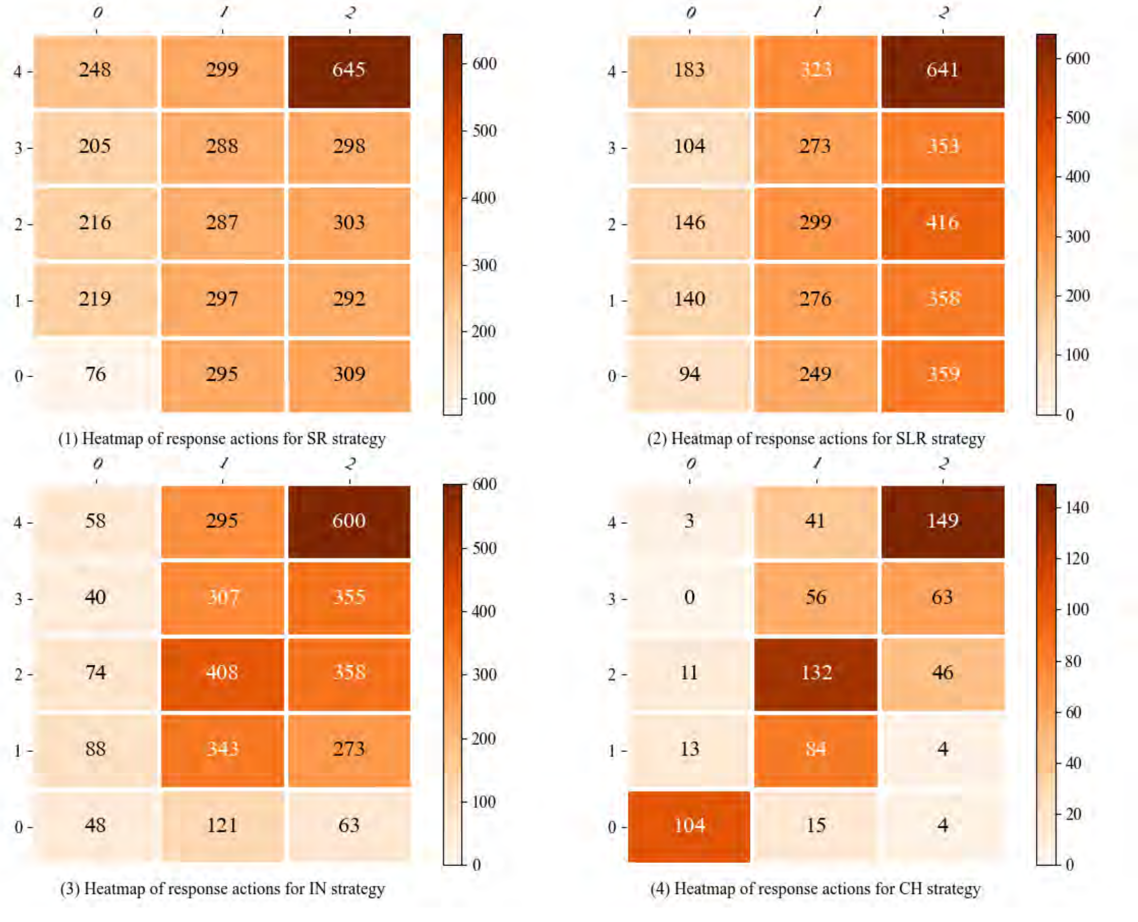
The heatmaps of response actions reflecting the ways the students designed experiments and applied scientific knowledge were analyzed. The horizontal coordinate of the heatmap represents "Towel Position," and the vertical coordinate represents "Water Level." The numbers in each grid in the heatmaps represent the number of clicks on each experimental combination. As shown in Figure 2, according to the shades of color in the graphics, the students who applied the SR strategy mainly focused on controlling variables "Towel Position 2", "Towel Position 1" and "No Towel", and especially collecting the data of combination of "Towel Position 2" and "Water Level 4", suggesting that they cared about all key variables when designing experiments and were more curious about the combination of "Towel Position 2, Water Level 4" for lowering temperature; the students who adopted the SLR strategy paid more attention to control variable "Towel Position 2", and especially collecting the data of combinations of "Towel Position 2, Water Level 4", and "Towel Position 2, Water Level 2", implying that they might have believed "Towel Position 2" could play a better role in decreasing temperature than other towel positions and cared about the roles of "Water Level 4" and "Water Level 2"; the students who used the IN strategy mainly focused on controlling variables "Towel Position 1" and "Towel Position 2", collecting experimental data of combinations of "Towel

https://doi.org/10.33225/jbse/23.22.929

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Position 2, Water Level 4", "Towel Position 1, Water Level 2", and "Towel Position 2, Water Level 2", and ignored the data of "Water Level 0" compared to the SR and SLR strategies, suggesting an incomplete and ambiguous experimental design; the students who adopted the CH strategy paid attention to the combinations of "Towel Position 2, Water Level 4", "Towel Position 1, Water Level 2", and "No Towel, Water Level 0", while the design of experiments was completely discontinuous and chaotic.

Figure 2
Heatmaps of Response Actions for the Strategy Groups



Additionally, to further explore the concrete and detailed trajectories of experimental design for solving scientific problems for each strategy group, the top two most representative response action sequences for each strategy were recorded. Table 5 presents the top two most representative response action sequences and the frequency, proportion and examples of these sequences within each strategy group. Due to the various operations in the task for problem solvers, there were 140 sorts of response sequences among 280 students for the SR strategy, 338 kinds of response sequences among 385 students for the SLR strategy, 409 kinds of response sequences among 478 students for the IN strategy, and 125 sorts of response sequences among 373 students for the CH strategy.

https://doi.org/10.33225/jbse/23.22.929

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Table 5
Top Two Most Representative Response Sequences Across Strategy Groups and Their Frequency and Proportion Within Each Strategy Group

Strategies	Representative Response Action Sequences	Frequency	Percentage (%)	Examples
Scientific and Rigorous (SR)	…‘[0, 0]’, ‘[0, 1]’, ‘[0, 2]’, ‘[0, 3]’, ‘[0, 4]’, ‘[1, 0]’, ‘[1, 1]’, ‘[1, 2]’, ‘[1, 3]’, ‘[1, 4]’, ‘[2, 0]’, ‘[2, 1]’, ‘[2, 2]’, ‘[2, 3]’, ‘[2, 4]’…	79	28.2	[‘[0, 0]’, ‘[0, 1]’, ‘[0, 2]’, ‘[0, 3]’, ‘[0, 4]’, ‘[1, 0]’, ‘[1, 1]’, ‘[1, 2]’, ‘[1, 3]’, ‘[1, 4]’, ‘[2, 0]’, ‘[2, 1]’, ‘[2, 2]’, ‘[2, 3]’, ‘[2, 4]’, ‘[2, 4]’, ‘[2, 4]’, ‘[2, 4]’, ‘[2, 2]’]
				[‘[0, 0]’, ‘[0, 1]’, ‘[0, 2]’, ‘[0, 3]’, ‘[0, 4]’, ‘[1, 0]’, ‘[1, 1]’, ‘[1, 2]’, ‘[1, 3]’, ‘[1, 4]’, ‘[2, 0]’, ‘[2, 2]’, ‘[2, 1]’, ‘[2, 3]’, ‘[2, 4]’]
	…‘[1, 0]’, ‘[1, 1]’, ‘[1, 2]’, ‘[1, 3]’, ‘[1, 4]’, ‘[2, 0]’, ‘[2, 1]’, ‘[2, 2]’, ‘[2, 3]’, ‘[2, 4]’…	72	25.7	[‘[1, 0]’, ‘[1, 1]’, ‘[1, 2]’, ‘[1, 3]’, ‘[1, 4]’, ‘[2, 0]’, ‘[2, 1]’, ‘[2, 2]’, ‘[2, 3]’, ‘[2, 4]’, ‘[2, 4]’, ‘[2, 4]’]
				[‘[0, 1]’, ‘[0, 2]’, ‘[0, 3]’, ‘[0, 4]’, ‘[1, 0]’, ‘[1, 1]’, ‘[1, 2]’, ‘[1, 3]’, ‘[1, 4]’, ‘[2, 0]’, ‘[2, 1]’, ‘[2, 2]’, ‘[2, 3]’, ‘[2, 4]’, ‘[2, 4]’, ‘[2, 4]’]
Scientific and Less Rigorous (SLR)	…‘[2, 0]’, ‘[2, 1]’, ‘[2, 2]’, ‘[2, 3]’, ‘[2, 4]’…	78	20.3	[‘[1, 1]’, ‘[1, 2]’, ‘[1, 3]’, ‘[1, 4]’, ‘[2, 0]’, ‘[2, 1]’, ‘[2, 2]’, ‘[2, 3]’, ‘[2, 4]’]
				[‘[0, 0]’, ‘[0, 0]’, ‘[0, 1]’, ‘[0, 3]’, ‘[0, 4]’, ‘[1, 4]’, ‘[2, 4]’, ‘[2, 4]’, ‘[0, 4]’, ‘[0, 4]’, ‘[2, 0]’, ‘[2, 1]’, ‘[2, 2]’, ‘[2, 3]’, ‘[2, 4]’, ‘[1, 4]’, ‘[1, 2]’, ‘[1, 1]’, ‘[1, 0]’]
	…‘[1, 0]’, ‘[1, 1]’, ‘[1, 2]’, ‘[1, 3]’, ‘[1, 4]’…	52	13.5	[‘[1, 0]’, ‘[1, 1]’, ‘[1, 2]’, ‘[1, 3]’, ‘[1, 4]’, ‘[2, 4]’, ‘[2, 4]’, ‘[2, 4]’]
				[‘[0, 0]’, ‘[1, 0]’, ‘[1, 1]’, ‘[1, 2]’, ‘[1, 3]’, ‘[1, 4]’, ‘[2, 1]’, ‘[2, 2]’, ‘[2, 4]’, ‘[2, 3]’, ‘[2, 3]’, ‘[2, 3]’]
Incomplete (IN)	…‘[1, 1]’, ‘[1, 2]’, ‘[1, 3]’, ‘[1, 4]’, ‘[2, 1]’, ‘[2, 2]’, ‘[2, 3]’, ‘[2, 4]’…	82	17.2	[‘[1, 1]’, ‘[1, 2]’, ‘[1, 3]’, ‘[1, 4]’, ‘[2, 1]’, ‘[2, 2]’, ‘[2, 3]’, ‘[2, 4]’, ‘[2, 4]’, ‘[2, 4]’]
				[‘[0, 0]’, ‘[1, 1]’, ‘[1, 2]’, ‘[1, 3]’, ‘[1, 4]’, ‘[2, 1]’, ‘[2, 2]’, ‘[2, 3]’, ‘[2, 4]’, ‘[2, 4]’, ‘[2, 4]’, ‘[2, 4]’]
	…‘[1, 1]’, ‘[1, 2]’, ‘[1, 3]’, ‘[1, 4]’…	48	10.0	[‘[1, 1]’, ‘[1, 2]’, ‘[1, 3]’, ‘[1, 4]’, ‘[2, 4]’]
				[‘[0, 0]’, ‘[0, 0]’, ‘[1, 1]’, ‘[1, 2]’, ‘[1, 3]’, ‘[1, 4]’, ‘[1, 4]’, ‘[2, 1]’, ‘[2, 2]’]
Chaotic (CH)	…‘[1, 2]’…	94	25.2	[‘[1, 0]’, ‘[1, 2]’, ‘[1, 2]’, ‘[1, 2]’]
				[‘[1, 2]’, ‘[1, 2]’, ‘[1, 2]’]
	…‘[2, 4]’…	81	21.7	[‘[0, 0]’, ‘[2, 4]’, ‘[2, 4]’, ‘[2, 4]’]
				[‘[1, 1]’, ‘[2, 2]’, ‘[2, 4]’, ‘[2, 4]’, ‘[2, 4]’]

First, among the students who applied the SR strategy, the most representative sequence was exhibited by 28.2% of the students and involved keeping the variable "Towel Position" (including No Towel, Towel Position 1, Towel Position 2, all three positions) constant and collecting five sets of "Water Level" data for each towel position. Another representative response sequence for the students who adopted the SR strategy involved keeping the variable "Towel Position" (including Towel Position 1 and Towel Position 2, two of the three positions) constant and collecting five sets of "Water Level" data for each towel position.

Second, among the students who used the SLR strategy, the most representative sequence presented in Table 5 involved keeping the variable "Towel Position" (including Towel Position 2 or Towel Position 1, only one of the three positions) constant and collecting five sets of "Water Level" data for each towel position. It is worth mentioning that the proportion of students who exhibited the actions of keeping the variable "Towel Position 2" constant and collecting five sets of "Water Level" data was more than that of the students who kept the variable "Towel Position 1" constant, which is consistent with the results of the heatmap of response actions.

Third, among the students who adopted the IN strategy, the most representative sequence exhibited by 17.2% of the students showed that although they were aware of keeping the variable "Towel Position" (including Towel Position 1 and Towel Position 2, two of the three positions) constant, they collected incomplete "Water Level" data

https://doi.org/10.33225/jbse/23.22.929

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

for each towel position. Another representative response sequence exhibited by 10.0% of the students involved keeping the variable "Towel Position" (including Towel Position 1, one of the three positions) constant but collecting incomplete data on "Water Level."

Last, for the CH strategy group, the top two representative sequences were exhibited by 25.2% and 21.7% of the students, respectively, suggesting that the students paid attention to the experimental combinations of "Towel Position 1, Water Level 2" and "Towel Position 2, Water Level 4". The findings are consistent with the results of the heatmap of response actions. The CH strategy group cannot design experiments to solve problems based on the principles of controlling variables. This group had no idea how to design experiments and merely solved problems irregularly and chaotically.

Differences in Response Time across Experimental Design Strategies

Table 6 shows that the students who adopted the SR strategy spent the most time-solving problems. In contrast, the students who used the CH strategy spent the least time responding to tasks, suggesting that the students who applied rigorous and scientific strategies needed more time to solve scientific problems.

Table 6
Descriptive Statistics of Total Response and Execution Time Across Strategy Groups

Measure	Scientific and Rigorous		Scientific and Less Rigorous		Incomplete		Chaotic	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Total response time (seconds)	532.26	309.46	428.85	270.13	348.09	273.64	231.78	228.54
Execution time (seconds)	71.66	46.12	59.75	50.02	41.98	34.07	21.04	34.01

Regarding the total response time, the results suggest that the means of response time for the four strategy groups were unequal according to a one-way ANOVA, F (3, 1471) = 71.603, $p < .05$. Pairwise comparisons of the means using Games-Howell tests indicated four significant comparisons: the mean time for the SR strategy (M = 532.26, SD = 309.46) was significantly ($p < .05$) longer than that for the SLR strategy (M = 428.85, SD = 270.13); the mean time for the SLR strategy (M = 428.85, SD = 270.13) was significantly ($p < .05$) longer than that for the IN strategy (M = 348.09, SD = 273.64); the mean time for the IN strategy (M = 348.09, SD = 273.64) was significantly ($p < .05$) longer than that for the CH strategy (M = 231.78, SD = 228.54); and the mean time for the CH strategy (M = 231.78, SD = 228.54) was significantly ($p < .05$) shorter than that for the SR strategy (M = 532.26, SD = 309.46).

Similar trends were found for the execution time. The results suggest that the means of execution time for the four strategy groups were unequal according to a one-way ANOVA, F (3, 1388) = 89.486, $p < .05$. Pairwise comparisons of the means using Games-Howell tests indicated four significant comparisons: the mean time for the SR strategy (M = 71.66, SD = 46.12) was significantly ($p < .05$) longer than that for the SLR strategy (M = 59.75, SD = 50.02); the mean time for the SLR strategy (M = 59.75, SD = 50.02) was significantly ($p < .05$) longer than that for the IN strategy (M = 41.98, SD = 34.07); the mean time for the IN strategy (M = 41.98, SD = 34.07) was significantly ($p < .05$) longer than that for the CH strategy (M = 21.04, SD = 34.01); and the mean time for the CH strategy (M = 21.04, SD = 34.01) was significantly ($p < .05$) shorter than that for the SR strategy (M = 71.66, SD = 46.12).

Discussion

Different Ways of Designing Experiments and Applying Scientific Knowledge Across Strategy Groups

The trajectories of the four strategies in this study reflected various ways of designing experiments and applying scientific knowledge when solving scientific problems. Regarding the ways of designing experiments, the SR group that achieved the highest total scores paid attention to all variables and knew how to scientifically control variables and

https://doi.org/10.33225/jbse/23.22.929

design rigorous experiments to explore the best conditions for lowering the temperature. The SLR group was also aware of controlling variables scientifically and mainly paid attention to the variables they cared about; therefore, the design of experiments was not rigorous and systematic enough. The IN group knew the variables needed to be controlled when designing experiments, while the experimental data for exploring the best condition for lowering temperature were not valid enough because the design of experiments was incomplete and nonrigorous. The CH group, which is also a low-performing group according to their total score, was not aware of controlling variables when designing experiments and designed experiments in a completely chaotic and irregular way. To explore the condition for the best refrigeration effect, students need to combine various variable conditions ("Towel Position" and "Water Level") to set up experimental combinations for comparison. The process of setting up experimental combinations requires a good deal of cognitive resources for controlling variables and designing experiments. The need for abundant cognitive resources to design experiments is consistent with the literature (Gong et al., 2020). The patterns of designing experiments to solve complex scientific problems suggested that successful problem solvers were generally inclined to explore all possible experimental combinations and design experiments scientifically and rigorously.

Regarding the application of scientific knowledge, the SR group paid more attention to the combination of "Towel Position 2, Water Level 4", and the SLR group cared more about the combinations of "Towel Position 2, Water Level 4" and "Towel Position 2, Water Level 2". Therefore, understanding the roles of water levels 2 and 4 was crucial for the students to design experiments. Water level 2 is related to the principle of evaporation and heat absorption. The towel absorbs water in the pot, and then the water evaporates and absorbs heat from the bowl. The process of evaporation and heat absorption plays a vital role in the refrigeration effect. Water level 4 is relevant to another scientific principle that heat transfer occurs when the water in the pot submerges the bowl. Due to heat transfer, water submerging in the bowl would transfer the heat of the water to the bowl to reduce the refrigeration effect. Understanding the aforementioned principles is useful for students to design experiments. The findings showed that many attempts were made to explore what would happen when water submerges the bowl (water level 4) and does not submerge the bowl (water level 2). The SR group students made many more attempts to discover the roles of water level 4 than level 2, implying that the SR group might be aware of the scientific principle of evaporation and heat absorption but not heat transfer. In contrast, the SLR group cared about water levels 4 and 2, suggesting that the SLR group might be confused about the principle of evaporation and heat absorption and the principle of heat transfer. The results indicated that the relevant scientific knowledge and principles in students' minds might affect the use of strategies and the ways they design experiments when solving scientific problems. Successful problem solvers have systematic and structured knowledge, while unsuccessful problem solvers tend to have less scientific and unorganized knowledge (Zajchowski & Martin, 1993). Our findings are consistent with some research showing that conceptual knowledge is basic for students to solve scientific problems and is a predictor of problem-solving competency (Friege & Lind, 2006).

Sufficient Time for Solving Complicated Scientific Problems

Results show that the students who used more scientific and rigorous strategies spent more time seeking solutions when solving scientific problems. Since the scientific problems that occur in scenario-based tasks and daily life are generally complicated and authentic, successful problem solvers usually need to spend more time making plans and finding solutions to better solve the problems in the processes of dealing with difficulties (Cowie, 2015; Gong et al., 2020). The previous findings are consistent with the total response time and execution time results for different strategy groups in this study. That means that to design a scientific and rigorous experiment to successfully solve a scientific problem, ample time might be used to reach the goal. When solving problems, while both successful and unsuccessful problem solvers consider a given problem situation in terms of the problem's objectives, successful problem solvers tend to spend sufficient time making assumptions and mapping the problem situation to an appropriate theoretical model by retrieving valid representations obtained from their larger and better-organized knowledge base (Yerushalmi & Eylon, 2015). Moreover, successful problem solvers may also spend more time than unsuccessful problem solvers reflecting on their former decisions and revising their choices accordingly (Yerushalmi & Eylon, 2015; Jiang et al., 2021).

An explanation for the relationships between time on task and task success may be provided by dual processing theory, which suggests a dynamic interaction of automatic and controlled mental processing (Schneider & Chein, 2003). Automatic processes that are fast do not require active control, whereas controlled processes that last longer than automatic processes require attentional control. Researchers argue that response time on a task is dependent on the relative degree of automatic versus controlled processing required by the task (Goldhammer et al., 2014). Due to the complex processes of problem solving containing knowledge acquisition and the application of this knowledge to

https://doi.org/10.33225/jbse/23.22.929

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

generate solutions, scientific problem solving must rely on controlled processing to a substantial degree, especially in some difficult tasks. Previous research has indicated that task difficulty affects the relationship between time on task and task success (Goldhammer et al., 2014). It can be assumed that easy tasks require essentially automatic processing to complete, whereas difficult tasks require a higher level of controlled processing. Although successful problem solvers proceed more quickly from information identification to integration, they need to invest sufficient time in seeking solutions rather than in low-level information processing. Accordingly, time spent on problem-solving tasks may increase with task difficulty.

Conclusions and Implications

The purpose of the present study was to explore the eleventh-grade Chinese students' problem-solving strategies and the differences among strategy groups by analyzing process data including sequences of students' response actions and response time. Results showed that students mainly adopted four strategies for designing experiments to solve complicated scientific problems. Successful problem solvers were generally inclined to explore all possibilities of experimental combinations and design experiments scientifically and rigorously based on the relevant scientific principles. There were significant differences in the response time used for solving problems among different strategy groups. The students who adopted more scientific and rigorous strategies spent more time designing plans and seeking solutions to solve complex scientific problems.

The results enrich the literature on using process data obtained from interactive items to address theoretical issues in educational assessment and provide actionable feedback for teaching and learning the skills required in scientific problem-solving tasks. First, students need to be guided and instructed to be aware of the principles of designing controlled experiments when solving scientific problems. Second, students' understanding of the principle of evaporation and heat absorption affected how they designed experiments to solve problems. Science teachers should help students construct a systematic and organized knowledge structure. Third, science teachers should give students sufficient time and ample opportunities to adequately participate in problem-solving activities and make elaborate plans for seeking solutions to cultivate students' scientific problem-solving ability. The length of time for carrying out problem-solving activities should be based on the students' academic performance (e.g., length of time to fully design and complete the experiment) and the difficulty of the tasks.

One limitation of this study is that the conclusions might not be applicable to all Chinese students since the data were collected from students in southern China through a convenience sampling technique with a limited number of sample students. Considerably more research should be conducted to discover patterns of students' problem-solving strategies in a large-scale assessment in other economic or cultural contexts. Additionally, more general conclusions about science could not be reached in this study since the analysis was performed only on one subtask isolated from the set of 17 subtasks in total and the subtask merely belonged to the domain of Physics. More subtasks covering all science-related domains should be analyzed in the future to complement and enrich the findings on students' problem-solving strategies. Moreover, the classification of strategies for solving scientific problems was mainly based on the analysis of process data reflecting students' response actions. In a follow-up study, it would be helpful to further explore the specific descriptions of each strategy group using qualitative analysis, such as think-aloud protocol analysis and interviews.

Acknowledgements

The authors thank Prof. Hongyun Liu for her help in providing expert reviews of the assessment and preparing the data. The authors thank Prof. Xingkai Luo and his team for their technical support. The authors also gratefully acknowledge the students, science teachers and schools that assisted with the data collection.

Declaration of Interest

The authors declare no competing interest.

References

Arslan, B., Jiang, Y., Keehner, M., Gong, T., & Katz, I. R. (2020). The effect of drag-and-drop item features on test-taker performance and response strategies. *Educational Measurement: Issues and Practice*, *39*(2), 96–106. https://doi.org/10.1111/emip.12326

Bergner, Y., & von Davier, A. A. (2019). Process data in NAEP: Past, present, and future. *Journal of Educational and Behavioral Statistics*, *44*(6), 706–732. https://doi.org/10.3102/1076998618784700

https://doi.org/10.33225/jbse/23.22.929

Camacho, M., & Good, R. (1989). Problem solving and chemical equilibrium: Successful versus unsuccessful performance. *Journal of Research in Science Teaching*, *26*, 251–272. https://doi.org/10.1002/tea.3660260306

Chang, C. Y., Barufaldi, J. P., Lin, M. C., & Chen, Y. C. (2007). Assessing tenth-grade students' problem-solving ability online in the area of earth sciences. *Computers in Human Behavior*, *23*(4), 1971–1981. https://doi.org/10.1016/j.chb.2006.02.014

Chang, C. Y., & Weng, Y. H. (2002). An exploratory study on students' problem-solving ability in earth science. *International Journal of Science Education*, *24*(5), 441–451. https://doi.org/10.1080/09500690110066502

Chi, M. T. H., De Leeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439–477. https://doi.org/10.1016/0364-0213(94)90016-7

Chung, G. K. W. K., & Baker, E. L. (2003). An exploratory study to examine the feasibility of measuring problem-solving processes using a click-through interface. *Journal of Technology, Learning, and Assessment*, *2*(2). https://ejournals.bc.edu/index.php/jtla/article/view/1662

Cowie, B. (2015). Problem solving in science, assessment of the ability to. In R. Gunstone (Eds.), *Encyclopedia of Science Education* (pp. 791–794). Springer Netherlands.

Eysenck, M. W., & Keane, M. (2000). *Cognitive psychology: A student's handbook*. Psychology Press.

Friege, G., & Lind, G. (2006). Types and qualities of knowledge and their relation to problem solving in physics. *International Journal of Science and Mathematics Education*, *4*, 437–465. https://doi.org/10.1007/s10763-005-9013-8

Gayford, C. (1989). A contribution to a methodology for teaching and assessment of group problem solving in biology among 15-year-old pupils. *Journal of Biological Education*, *23*, 193–197. https://doi.org/10.1080/00219266.1989.9655067

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology*, *106*(3), 608–626. https://doi.org/10.1037/a0034716

Gong, T., Shuai, L., Arslan, B., & Jiang, Y. (2020). Process based analysis on scientific inquiry tasks using large-scale national assessment dataset. In Proceedings of the 13th International Conference on Educational Data Mining (EDM 2020), Virtual Event.

Greiff, S., & Neubert, J. C. (2014). On the relation of complex problem solving, personality, fluid intelligence, and academic achievement. *Learning and Individual Differences*, *36*, 37–48. https://doi.org/10.1016/j.lindif.2014.08.003

Greiff, S., Wüstenberg, S., & Avvisati, F. (2015). Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving. *Computers & Education*, *91*, 92–105. https://doi.org/10.1016/j.compedu.2015.10.018

Greiff, S., Wüstenberg, S., Csapó, B., Demetriou, A., Hautamäki, J., Graesser, A. C., Martin, R. (2014). Domain-general problem solving skills and education in the 21st century. *Educational Research Review*, *13*, 74–83. https://doi.org/10.1016/j.edurev.2014.10.002

Guo, H., Rios, J. A., Haberman, S., Liu, O. L., Wang, J., & Paek, I. (2016). A new procedure for detection of students' rapid guessing responses using response time. *Applied Measurement in Education*, *29*(3), 173–183. https://doi.org/10.1080/08957347.2016.1171766

Jiang, Y., Gong, T., Saldivia, L. E., Cayton-Hodges, G., & Agard, C. (2021). Using process data to understand problem-solving strategies and processes for drag-and-drop items in a large-scale mathematics assessment. *Large-scale Assessments in Education*, *9,* Article 2. https://doi.org/10.1186/s40536-021-00095-4

Karadakal, N. V., Goud, N., & Thomas, P. (2015). Impact of leadership role perspective on conflict resolution styles - a study on small and medium sized entrepreneurs of Karnataka State in India. *Journal of Global Entrepreneurship Research*, *5,* Article 4. https://doi.org/10.1186/s40497-015-0019-6

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*(1), 1–48. https://doi.org/10.1207/s15516709cog1201_1

Lavoie, D. R. (1993). The development, theory, and application of a cognitive-network model of prediction problem solving in biology. *Journal of Research in Science Teaching*, *30*(7), 767–786. https://doi.org/10.1002/tea.3660300713

Lazakidou, G., & Retalis, S. (2010). Using computer supported collaborative learning strategies for helping students acquire self-regulated problem-solving skills in mathematics. *Computers & Education*, *54*(1), 3–13. https://doi.org/10.1016/j.compedu.2009.02.020

Lee, Y. H., & Jia, Y. (2014). Using response time to investigate students' test-taking behaviors in a NAEP computer-based study. *Large-Scale Assessments in Education*, *2*(8). https://doi.org/10.1186/s40536-014-0008-1

Mayer, R. E. (1992). *Thinking, problem solving, cognition* (2nd ed.). Freeman.

McIntosh, T. C. (1995). Problem-solving processes. *Science Teacher*, *62*(4), 16–19.

National Research Council (1996). *National science education standards*. National Academy Press.

National Research Council (2010). *Exploring the intersection of science education and 21st century skills: a workshop summary*. National Academy Press.

National Research Council (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. The National Academies Press.

OECD (2013a). *The PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. PISA, OECD Publishing.

Polya, G. (1957). *How to solve it* (2nd ed.). Lawrence Erlbaum.

Provasnik, S. (2021). Process data, the new frontier for assessment development: Rich new soil or a quixotic quest? *Large-scale Assessments in Education*, *9*(1). https://doi.org/10.1186/s40536-020-00092-z

Reid, N., & Yang, M. J. (2002). Open-ended problem solving in school chemistry: A preliminary investigation. *International Journal of Science Education*, *24*(12), 1313–1332. https://doi.org/10.1080/09500690210163189

Robinson, W., & Niaz, M. (1991). Performance based on instruction by lecture or by interaction and its relationship to cognitive variable. *International Journal of Science Education*, *13*(2), 203–215. https://doi.org/10.1080/0950069910130207

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Ross, J. A., & Maynes, F. J. (1983). Development of a test of experimental problem-solving skills. *Journal of Research in Science Teaching*, *20*(1), 63–75. https://doi.org/10.1002/tea.3660200107

Schauble, L. (2003). Scientific thinking: More on what develops. *Human Development*, *46*(2-3), 155–160. https://doi.org/10.1159/000068588

Schneider, W., & Chein, J. M. (2003). Controlled & automatic processing: Behavior, theory, and biological mechanisms. *Cognitive Science*, *27*(3), 525–559. https://doi.org/10.1207/s15516709cog2703_8

Sternberg, R. J. (1995). Expertise in complex problem solving: A comparison of alternative concepts. In P. A. Frensch, & J. Funke (Eds.), *Complex problem solving. The European perspective* (pp. 295–321). Hillsdale, Lawrence Erlbaum.

Sugrue, B. (1995). A theory-based framework for assessing domain-specific problem-solving ability. *Educational Measurement: Issues and Practice*, *14*(3), 29–35. https://doi.org/10.1111/j.1745-3992.1995.tb00865.x

Van der Linden, W. J. (2008). Using response times for item selection in adaptive testing. *Journal of Educational and Behavioral Statistics*, *33*(1), 5–20. https://doi.org/10.3102/1076998607302626

Vendlinski, T. & Stevens, R. (2002). Assessing student problem-solving skills with complex computer-based tasks. *Journal of Technology, Learning, and Assessment*, *1*(3). Retrieved from https://ejournals.bc.edu/index.php/jtla/article/view/1669

Yerushalmi, E., & Eylon, B. S. (2015). Problem solving in science learning. In R. Gunstone (Eds.), *Encyclopedia of science education* (pp. 786–790). Springer Netherlands.

Zajchowski, R. & Martin, J. (1993). Differences in the problem solving of stronger and weaker novices in physics: Knowledge, strategies, or knowledge structure. *Journal of Research in Science Teaching*, *30*(5), 459–470. https://doi.org/10.1002/tea.3660300505

Received: *June 30, 2023* Revised: *August 09, 2023* Accepted: *October 01, 2023*

Cite as: Zhao, P., Chang, C.-Y., Shao, Y., Liu, Z., Zhou, H., & Liu, J. (2023). Utilization of process data in China: Exploring students' problem-solving strategies in computer-based science assessment featuring interactive tasks. *Journal of Baltic Science Education*, *22*(5), 929-944. https://doi.org/10.33225/jbse/23.22.929

Pingping Zhao	PhD, Assistant Professor, College of Education for the Future, Beijing Normal University at Zhuhai, No. 18 Tangjiawan Jinfeng Road, Xiangzhou District, Zhuhai 519085, China. E-mail: pinggsshyxqa@126.com ORCID: https://orcid.org/0000-0002-6920-6053
Chun-Yen Chang	PhD, Chair Professor, Graduate Institute of Science Education, National Taiwan Normal University, 88, Section 4, Ting-Jou Rd., Taipei City 116, Taiwan. E-mail: changcy@ntnu.edu.tw ORCID: https://orcid.org/0000-0003-2373-2004
Yueyang Shao	PhD Candidate, Collaborative Innovation Center of Assessment Toward Basic Education Quality, Beijing Normal University, No. 19, XinJieKouWai St., HaiDian District, Beijing 100875, China. E-mail: a1288370@gmail.com
Zhi Liu	PhD Candidate, Collaborative Innovation Center of Assessment Toward Basic Education Quality, Beijing Normal University, No. 19, XinJieKouWai St., HaiDian District, Beijing 100875, China. E-mail: liuzhijobs@163.com
Hao Zhou	PhD, Postdoctoral Researcher, Collaborative Innovation Center of Assessment Toward Basic Education Quality, Beijing Normal University No. 19, XinJieKouWai St., HaiDian District, Beijing 100875, China. E-mail: hao.zhou@bnu.edu.cn ORCID: https://orcid.org/0000-0002-3232-657X
Jian Liu *(Corresponding author)*	Bachelor, Full Professor, Collaborative Innovation Center of Assessment Toward Basic Education Quality, Beijing Normal University, No. 19, XinJieKouWai St., HaiDian District, Beijing 100875, China. E-mail: professorliu9506@126.com ORCID: https://orcid.org/0000-0001-7501-6881