

Eye-movement study of high- and low-prior-knowledge students' scientific argumentations with multiple representations

Chao-Jung Wu[ⓧ] and Chia-Yu Liu^{ⓧ*}

*Department of Educational Psychology & Counseling,
Institute for Research Excellence in Learning Sciences,
National Taiwan Normal University, Taipei 10644, Taiwan*



(Received 23 October 2020; accepted 2 March 2021; published 16 April 2021)

The purpose of this study was to explore how students with high- and low-prior-knowledge employed multiple representations in argumentation evaluation and generation tasks. The argumentation performance and eye-movement behaviors of 96 college students in these tasks were investigated. The number of participants who proposed complex argumentation levels and the argumentation accuracy was higher in the high-prior-knowledge group than in the low-prior-knowledge group. Moreover, the high-prior-knowledge group demonstrated greater eye-movement transitions between representations compared with the low-prior-knowledge group. Both groups had greater transitions in the generation task than the evaluation tasks. The high-prior-knowledge group distributed attention to representations with more flexibility, revealing that they were more aware of the task requirements and more able to employ multiple representations for arguments. In the argumentation evaluation tasks, the high-prior-knowledge group performed referencing behaviors in the reading sequences between representation text and equation and between representation table and figure, whereas the low-prior-knowledge group was inclined to look back and forth between representation text and table. In the argumentation generation task, the two groups displayed similar reading sequences. It indicated that learners with higher knowledge may perceive the similarity between homogeneous representations and constrained interpretations of the complex representations by using easier representations, or further integrated representations to achieve deeper understanding, which then improved their argumentation performance. The implications of instructions for improving learners' argumentation skills in a multirepresentational display are discussed.

DOI: [10.1103/PhysRevPhysEducRes.17.010125](https://doi.org/10.1103/PhysRevPhysEducRes.17.010125)

I. INTRODUCTION

Science is not a collection of unproblematic facts about the world. It may generate controversies between scientists, and puzzling events can take place. Scientific argumentation is the central feature [1] of the resolution of scientific controversies. Argumentation ability should be developed in young people in any scientific education program [2]. However, argumentation activity is seldom included as a regular science class [3,4], and the OECD [5] has indicated that numerous students have difficulty identifying complicated arguments, evaluating the quality of arguments using evidence, and applying related knowledge to real-life contexts. At least two types of scientific argumentation should be considered. One is to evaluate the argument's credibility using evidence and another challenging one is to

further generate scientific argumentations by combining evidence and scientific theory. Several studies have indicated that the argumentation performance of individuals is improved by more relevant knowledge and less demanding tasks [6–8]. To confirm the validity of our materials, the first research question was as follows: Are argumentation performances better among high-prior-knowledge learners and during less demanding tasks?

Furthermore, physics education often contains multiple forms of representation (e.g., text, figure, equation, and table), requiring learners to select and integrate information and then construct their arguments. Understanding multiple representations is a complex task. For kinematics, studies over the past three decades have revealed students' difficulty in learning from multiple representations [9]. Numerous studies in physics education have focused on promoting learners' ability to use multiple representations [10,11], and several studies have used eye trackers to investigate individuals' mental processing [11–13]. However, few studies have examined how learners apply multiple representations for scientific argumentation. Clarifying how approaches to allocating attention to representations for argumentation differ between learners

*leave1756@gmail.com

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

based on their levels of knowledge allows us to establish a solid foundation for future interventions using multiple representations. Several researchers have adopted introspective reports or verbal protocol methods to examine individuals' argumentation performance (e.g., Ref. [14]). We argue that recording learners' eye movements as they produce arguments is critical because reasoning behaviors are not necessarily conscious and may be difficult to observe. Eye tracking reveals how learners process information during cognitive activities by recording which information is considered and how deeply it is processed [12,15,16]. Although the relationship between gaze direction and attention location is not exact, it is believed to be highly positive, especially during challenging tasks, such as reasoning (e.g., Ref. [7]). For example, reading time reflects cognitive effort while considering certain representations; transitions between representations reflect how learners integrate information; reading sequences reveal learners' patterns when integrating multiple representations. Therefore, the second research question is as follows: What eye movements did learners with high- and low-prior knowledge exhibit in a multirepresentational display when engaging in argumentation tasks?

A. Learning from multiple representations

Physics education research studies over the past few decades have demonstrated that multiple representations play a critical role in successful physical learning [10,11,17]. Ainsworth's [18] framework proposed three functions of multiple representations that assist learning, which were complementary, constraining, and constructing. The complementary function refers to the use of various representations that provide complementary information or support distinctive inferences. The constraining function refers to when a familiar or easier representation is used to constrain the interpretation of the second abstract or complex one, thus supporting understanding. The constructing function refers to the adoption of multiple representations to achieve a deeper understanding of a concept. Based on Ainsworth's framework, Ott, Brünken, Vogel, and Malone [19] explored how the combinations of type code (i.e., symbolic and analog) affected mathematical problem solving. Symbolic representations are characterized as being linear and arbitrary, such as texts and equations, whereas analog representations consist of icons, such as pictures. Ott *et al.* observed that the combination of multiple homogeneous representations (i.e., text and equation) also contained a multimedia effect; this was similarly helpful to the combination of heterogeneous representations (i.e., text and graphic). More transitions were observed between multiple heterogeneous representations than multiple homogeneous representations. Furthermore, they confirmed that text was the reference representation that was most frequently inspected in their experiments.

Studies have reported that an individual's top-down knowledge affected their representation integration in various domains [11,17,20]. Kohl and Finkelstein [17] interviewed expert and novice physicists while they solved multiple representation problems. The expert solved problems more efficiently than novices and transferred their attention more rapidly between the available representations, indicating higher representational flexibility. Kozma [20] also adopted the expert-novice paradigm and determined that expert chemists could flexibly transfer their attention across multiple representations to understand the critical features of each topic, whereas chemistry students had difficulty coordinating different representations, and their understanding was often constrained by the surface features of a single representation. Klein *et al.* [11] also reported significant differences between physics majors who scored highest and lowest in the pretest. The high-scoring students demonstrated more saccades in the related direction of the arrow field and more systematic eye movements, indicating that individuals with higher levels of relevant knowledge employ suitable representations more readily.

Furthermore, task type also affects individuals' behaviors in representations selection and integration [12,21]. Since Curcio's work [22], several researchers have proposed similar tasks corresponding to three levels of graph comprehension [12,21–23]: the elementary level—extracting data from representations (i.e., reading the data), the intermediate level—identifying relationships among the data based on representations (i.e., reading between the data), and the advanced level—using representations to make predictions and answer questions (i.e., reading beyond the data). These researchers have all reaffirmed the need to consider all three types of tasks. For example, Meyer *et al.* [21] developed various information extraction tasks (e.g., reading exact values, comparing values, and identifying trends) and investigated learners' problem-solving performance using multiple representations (i.e., line graphs, bar graphs, and tables). The results indicated that line graphs were identified more rapidly than bar graphs for the task of identifying the trend of a data series, whereas the opposite tendency was observed for the task of reading exact values for a single point. Moreover, Schnotz *et al.* [12] developed combinations of texts and pictures and series of comprehension test items with various difficulty levels (i.e., specific information, simple relations, and complex relations), to explore the reading strategies used by fifth graders. They determined that participants' average number of fixations and accumulated fixation duration on pictures increased as the difficulty of items increased, whereas the trend was opposite for text.

Studies have emphasized that both individual knowledge and task type contribute to learners' representational skills [12,13,24]. Kriz and Hegarty [24] proposed a model of the comprehension process and evaluated the effects of

top-down processes (i.e., domain knowledge) and bottom-up processes (i.e., design of material) on readers' understanding and knowledge application. The results revealed that the design of learning materials may only be beneficial when an individual had adequate knowledge, indicating that learning involves the complex interplay. Based on the top-down and bottom-up model, Madsen *et al.* [13] examined the effects of physics knowledge and the perceptual saliency of the material on individuals' attention. They determined that participants who responded correctly spent more time inspecting the task-relevant areas of a diagram, and those who responded incorrectly spent more time viewing the novicelike answers. However, they also determined that the time spent in perceptual salient areas was similar for both correct and incorrect solvers, demonstrating that the bottom-up processes had limited effects on individual performances. Furthermore, Schnotz *et al.* [12] investigated lower- and higher-tier students' eye-movement behaviors under problem conditions for three difficulty levels. The results indicated that the higher-tier students' fixation frequency and accumulated fixation duration on pictures exhibited a considerable increase as the items became more difficult, whereas the trend was not observed among lower-tier students, indicating that individuals with more knowledge and learning skills were able to sense item difficulty and solve related problems.

B. Argumentation in science education

Argumentation is a form of meaningful discourse; its related aim is inducing deeper understanding and knowledge [25]. Duschl and Osborne [25] further suggested that both the philosophical and cognitive foundations of argumentation shape the meaning of argumentation in science education. The philosophical foundation emphasized that science often involves disputes and controversies rather than the accumulation of unproblematic facts. The cognitive foundation focused on how learners externalize their thinking from the *intrapsychological* state to the *interpsychological* argument [2].

Studies have demonstrated that providing familiar discussion topics [6] and highly contextualized tasks that engage students' prior knowledge [26] can improve argumentation performance. Moreover, the participants with relevant prior knowledge often had better argumentation performance and profited more from training [7,8,27]. Inglis and Alcock [7] determined that undergraduate students employed more time on the surface features of arguments instead of the critical logical structure. By contrast, mathematicians generally viewed the consecutive lines of proofs in a back and forth manner, indicating that they spent more cognitive effort on inferring implicit warrants. These studies revealed that individuals' understanding of related theories, models, laws, or conceptions affected how they supported, challenged, and proposed arguments.

Toulmin's argumentation pattern (TAP) [28] provided an influential framework for understanding the elements inferencing from raw data to reach a conclusion. The structure of TAP [28] has been widely used to examine the quality of individuals' argumentation performance [2,14,26]. According to TAP, full-fledged arguments contain six key components, comprising a claim, data, a warrant, a backing, a rebuttal, and a qualifier. Studies have reported that participants demonstrated better argumentation performance when the structure of the argument was simple (i.e., contains limited elements) and typical (i.e., claim-first argument; [8]), or the form of the argument was familiar to participants [6]. These findings indicated that the demand and type of argumentation tasks have various effects on the performance of individuals.

However, the definitions of the six components of TAP could be ambiguous [2,14,26]. Cavlazoglu and Stuessy [14] modified the scheme by excluding rebuttals, which should only occur in dialog discourse. Then, they examined the monologic discourse argumentation of science teachers and presented their concept maps in a teacher workshop. Erduran *et al.* [2] also adapted TAP to examine the quality of eighth-graders' science discourses. Studies have generally adjusted their classification schemas according to research design, participant characteristics, and argumentation type to more accurately examine individual performance.

C. Eye-movement indicators for investigating multiple representations

The eye-tracking methodology has been utilized to explore reading processes for several decades, and numerous studies in science education have adopted eye-tracking methodology to reveal the cognitive processes of individuals with various characteristics [12,16,27,29]. Studies have demonstrated that several eye-movement indicators could be used to investigate reading multiple representations [12,15,16,27,29]. First, total fixation duration (TFD) measures the sum of the duration of all fixations within an area of interest (AOI) and reflects the extent of cognitive effort for processing reading material. A longer TFD reflects greater cognitive effort required to understand the information [12,27,29]. Second, transition numbers between representations measure the duration of the eye-fixation movement from one representation to another. This measure reflects the integration of information among representations [12]. For example, Schnotz *et al.* [12] calculated the eye-movement transitions between text, picture, and item areas and determined that higher-tier students made more transitions between these areas compared with lower-tier students. Third, the proportion of fixation duration measures the fixation duration on specific AOIs divided by the TFD during the task [16,29]. For example, Stieff *et al.* [16] determined that students' proportions of fixation duration on graphs and models

were higher than for equations, indicating that the students may have lacked an understanding of more abstract representations and thus paid much less attention to the equations. Fourth, reading sequence can be calculated with the transition fixations from target AOIs to the next AOIs, reflecting the pattern of integrating multiple representations [15,16]. For example, Jian [15] adopted a sequential analysis and determined that most young readers referred from the illustrations to the texts unidirectionally, indicating that they generally focused on a single representation rather than multiple representations.

Although several argumentation-focused studies have adopted an eye-tracking methodology [7,27,29–34], to our knowledge, no study has investigated individuals' eye movement while selecting representations, integrating information, and evaluating and generating arguments. Eye trackers have been used to investigate individuals' reading behavior and the effects of reading behavior on argumentation performance [29,32], to record eye movement as immediate feedback for argumentation tasks [27], to explore differences in eye movement during argument reading based on an individual's characteristics [7,30], and to investigate the effects of argument structure and strength on learners' eye movement and attitude [31,33,34]. Unsurprisingly, argumentation-focused studies have analyzed similar eye movement indicators to investigate individuals' argumentation behavior. First, TFD has been used in all related studies as an indicator of an individual's cognitive attention [7,27,29–34]. Second, transition numbers between AOIs have also been used. For example, Inglis and Alcock [7] calculated line-transition matrices and found that mathematicians made far more between-line saccades than undergraduates, indicating the cognitive effort mathematicians exerted while inferring implicit warrants. Third, the proportion of fixation duration has also been investigated. In one study, Yang [29] divided the TFD in a text area by TFD on the whole page to calculate the proportion of fixation duration and determined that the participants did not allocate equal attention to text areas containing various elements of scientific argumentation.

D. The current study

Different from previous studies, an eye-tracking methodology was adopted to investigate the effect of an individual's prior knowledge and the task type on online argumentation performance and eye-movement behaviors. The first research question in this study assessed individuals' scientific argumentation performance. Studies by Grooms *et al.* [6], Inglis and Alcock [7], von der Mühlen *et al.* [8], and Tsai *et al.* [27] have demonstrated that both knowledge and task type affect the argumentation performance of learners. Individuals with more relevant knowledge should be able to propose higher-quality and higher-accuracy arguments than individuals with limited

knowledge. Moreover, if the tasks had lower demands (i.e., argumentation evaluation tasks), individuals should also produce higher-quality and higher-accuracy arguments (Hypothesis 1-1) (Hypothesis 1-2).

The second research question concerned individuals' eye-movement behaviors. Klein *et al.* [11], Kohl and Finkelstein [17], and Kozma [20] have reported that participants who had more top-down knowledge had a larger number of transitions between representations and had higher sensitivity to the task demand. Therefore, we predicted that individuals with more prior knowledge would exhibit a larger number of transitions between representations, especially during the more demanding task (i.e., argumentation generation task) (Hypothesis 2-1). On the basis of the findings by Schnotz *et al.* [12], Madsen *et al.* [13], and Kriz and Hegarty [24], we also predicted that individuals with higher prior knowledge would have a higher awareness of the task demands and devote more attention to the appropriate representations to evaluate (Hypothesis 2-2) and propose scientific arguments (Hypothesis 2-3). The correspondence between each task and the appropriate representation is reported in Sec. II B.

Hypothesis 1-1 Argument quality will be higher in the high-prior-knowledge group, especially for the argumentation evaluation tasks.

Hypothesis 1-2 Argument accuracy will be higher in the high-prior-knowledge group, especially for the argumentation evaluation tasks.

Hypothesis 2-1 The high-prior-knowledge group will display more transitions between representations, especially for the argumentation generation task.

Hypothesis 2-2 The high-prior-knowledge group will distribute more attention to the appropriate representations in the argumentation evaluation tasks.

Hypothesis 2-3 The high-prior-knowledge group will distribute more attention to the appropriate representations in the argumentation generation task.

A learner's reading sequence reveals their patterns in integrating multiple representations to argumentation, which should be used to develop instructions. Therefore, we further analyzed the eye movements of the two groups to investigate one exploratory research question.

Exploratory research question: What reading sequences do participants with low and high levels of prior knowledge exhibit during argumentation tasks?

II. METHOD

A. Participants

A total of 196 college students in Taiwan were recruited to participate in the prior knowledge test. The participants signed up through an independent website or experimenter recruitment emails. Only participants who scored in the top and bottom third of all participants were invited to participate in the following eye-movement experiment. The participants

were assigned to two groups: the high-prior-knowledge ($n = 49$) and low-prior-knowledge ($n = 47$) groups. All participants had normal or corrected-to-normal vision, and their average age was 21.87 yr ($SD = 2.56$).

B. Materials

1. Prior knowledge test

The Chinese version of the Test of Understanding Graphs in Kinematics (CTUG-K), translated from Beicher [35], was adapted to evaluate participants' prior knowledge of kinematics. The original test consisted of 21 items and assessed 7 dimensions. The test focused on examining kinematic graph interpretation skills, such as determining displacement from a velocity–time graph, selecting a textual description from a kinematic graph, and selecting the corresponding graph from a textual motion description. To discriminate between students with different levels of prior knowledge, 17 of the more difficult items were selected and tested in 94 university students. Afterward, we removed the items with lower difficulty, thus 10 items and the original 7 dimensions were retained in the CTUG-K, with possible scores ranging from 0 to 10. Among the 96 participants included in the formal study, the range of performance accuracy was 0.29–0.66 ($Md = 0.54$), item difficulty was 0.30–0.67 ($Md = 0.50$), item discrimination was 0.47–0.77 ($Md = 0.61$), and $KR_{20} = 0.86$, indicating that CTUG-K exhibited satisfactory measurement properties.

2. Tasks

The task scheme proposed by Curcio [22] was adapted into two types of tasks: argumentation evaluation and argumentation generation. Three argumentation evaluation tasks and one argumentation generation task were developed (e.g., Fig. 1). The topic and data concerned a weather balloon ascent in Coffeyville, Kansas, United States, in 1991, retrieved from a study by Erickson [36].

The argumentation evaluation tasks consisted of an identification task, a correlation task, and a prediction task, which corresponded to Curcio [22]'s three levels of graph comprehension. The participants were asked to evaluate whether the task's statement was true and justify their statement using the representations. The statement of the identification task was "Thor believed that the acceleration of the weather balloon was initially 0.05 m/s^2 , followed by a gradual decrease to 0 m/s^2 ." The table was appropriate for this task because participants should support their arguments by selecting the exact value of a certain point. The statement of the correlation task was "Hulk proposed that the weather balloon was moving at a constant speed over time." The figure was appropriate for this task because it visually represented the relationships between two variables. The statement of the prediction task was "Spiderman thought that the weather balloon would ascend to 20 000 meters in 4000 seconds." The equation was appropriate for this task because the participants could use it to estimate data trends.

Question: Spiderman thought that the weather balloon would ascend to 20,000 meters in 4,000 seconds. Please indicate that whether this statement is true and support your answer with the data.

Variable	Height	Time	Speed	Acceleration
Abbreviation	d	t	v	a
Unit	m	s	m/s	m/s*s
Data of the weather balloon	82.00	5.20	4.15	0.05
	271.00	18.40	4.56	0.05
	313.00	26.20	4.78	0.04
	400.00	42.90	4.94	0.02
	575.00	77.60	4.98	0.01
	664.00	95.20	4.99	0.01
	846.00	130.70	5.03	0.01
	1031.00	170.20	4.95	0.01
	1611.00	288.40	4.93	0.00
	2018.00	377.60	4.85	0.00
	2445.00	482.50	4.68	0.00
	2894.00	584.40	4.63	0.00
	3126.00	641.80	4.58	0.00
	3859.00	838.70	4.38	0.00
	5510.00	1246.50	4.27	0.00
	6456.00	1478.80	4.24	0.00
	8684.00	2035.30	4.18	0.00
	9566.00	2252.10	4.17	0.00
10547.00	2782.50	4.17	0.00	
12318.00	2895.90	4.18	0.00	

$$d = f(t) = 4t + 50 \quad \text{if } t > 1000$$

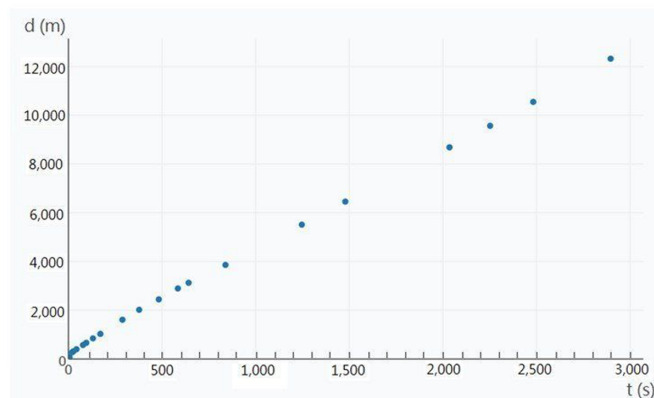


FIG. 1. The prediction task.

The argumentation generation task required that participants provide claims and support them using representations. The question was “Please observe this phenomenon, tell me about your claims in two to four sentences, and provide evidence for your claims.” The table was especially appropriate in this task since it contained the most information for participants to provide free-form answers. The four tasks contained the same table, but the figures and equations were different to correspond to each statement. For example, in the prediction task (Fig. 1), a distance–time graph and equation were presented because the statement involved distance and time variables. Appendix B displays the other tasks. Furthermore, a practice task on the topic of free fall was also developed to assist participants in understanding the experiment.

3. Apparatus

Eye movements were recorded using a Tobii X3-120 eye tracker at a sampling rate of 120 Hz. Participants were seated approximately 70 cm in front of the monitor with their heads stabilized on a forehead rest. The monitor’s diagonal viewing size was 27 inches with the 1920 × 1080-pixel resolution. The data were recorded using Tobii-Studio software (version 3.4.8).

4. Procedure

We collected data during two sessions. In the first session, 196 participants were collectively administered the CTUG-K for up to 10 min. The participants who scored in the top and bottom third of all participants were invited to the next session and assigned to two groups: the high-prior-knowledge ($n = 49$) and low-prior-knowledge ($n = 47$) groups.

The second session was the eye-tracking experiment with a digital voice recorder; it was tested individually in a quiet room. After informed consent was obtained from participants, they were shown the practice task and a worksheet to learn to express their arguments. The following statements were announced to the participants: “There is no correct answer to these questions,” “It is a set of authentic data, so the simplified equations can only demonstrate the trend,” and “The values in the table are rounded to the second decimal place.” For each participant, eye movements were calibrated using a 9-point calibration before each task. The participant was asked to voice their

arguments while they read the practice tasks on the screen and then click the mouse to finish the task. The process of formal tasks was identical to the practice task. The tasks’ sequence was the same for each participant. Participants read the material and completed the tasks at their own pace. The average time required for participants to finish reading the material and complete a task was approximately 8.24 min (SD = 3.65). On average, participants completed the procedure in approximately 30 min.

5. Data analysis

The aforementioned four eye-movement indicators (see the eye-movement indicators for investigating multiple representations section) were used in this study. First, the TFD for each task was calculated. Second, the *numbers of transitions* between the three representations were examined. Third, each task page was divided into several AOIs: the question stem, table, figure, and equation. The *proportion of fixation durations* was derived from the fixation duration on a certain AOI divided by the fixation duration on all AOIs during a task. Fourth, the *reading sequence* was assessed using sequential analysis [15], which was performed using a series of matrix multiplications.

The participants’ argumentation performance was analyzed for quality and accuracy. Our instrument for analyzing the quality was based on the argumentation model by Toulmin [28] and modified classification schema by Erduran *et al.* [2] and Cavlazoglu and Stuessy [14]. The modified schema consisted of five levels from 1 to 5, as illustrated in Table I. However, several levels contained fewer than five participants; thus a chi-square analysis could not be performed. Therefore, we combined the first two levels into a simple level and the others into a complex level. For example, arguments used to solve the prediction task (Fig. 1) included “I suggest the statement is wrong (CLAIM) since there is a disparity between the value in the statement and the value calculated with the equation (DATA),” argued by a participant in the simple level, and “The balloon should not ascend to 20 000 meters in 4000 seconds...maybe only 18 000 meters (CLAIM + DATA). From the time and height data, the balloon could exceed 20 000 meters as long as it does not rupture (QUALIFIER),” argued by a participant in the complex level. The arguments were further analyzed and discussed in another study. For the argument

TABLE I. Argumentation framework for assessing argumentation quality.

Levels	Components	
Simple	Level 1	claim or data or warrant
	Level 2	claim+(data or warrant)
Complex	Level 3	claim+(data or warrant)+(backing or qualifier or rebuttal)
	Level 4	claim+(data or warrant)+(backing and qualifier) or (backing and rebuttal) or (qualifier and rebuttal)
	Level 5	claim+(data or warrant)+backing+qualifier+rebuttal

level, each participant's highest argumentation level among the three argumentation evaluation tasks was used as their argumentation level in the argumentation evaluation tasks.

The accuracy of the participants' argumentation performance in each task was examined, and incorrect and correct arguments were scored 0 and 1, respectively. The accuracy in the argumentation evaluation tasks was calculated by averaging the accuracy in the identification, correlation, and prediction tasks. The first researcher scored 384 responses on the 4 tasks and another researcher reviewed 64 of the responses to assess the reliability of coding. The interrater reliability for argumentation level ($\kappa = 0.99$, $p < 0.001$) and accuracy ($\kappa = 0.92$, $p < 0.001$) were almost perfect.

III. RESULTS

A. Argumentation performance

To verify hypothesis 1-1 and 1-2, we performed a multivariate chi-squared analysis for argument level and a mixed ANOVA on argument accuracy to confirm whether the argument performances will be better for individuals with more prior knowledge and during less demanding tasks. The results of multivariate chi-squared analysis demonstrated that the variables group, task type, and argumentation level were significantly related ($G^2 = 22.47$, $p < 0.001$). When further investigating these relationships, we determined that the variables of group and argumentation level were significantly related ($G^2 = 20.86$, $p < 0.001$). The number of participants who displayed a complex argumentation level was significantly higher in the high-prior-knowledge group than that in the low-prior-knowledge group, and the number of participants who exhibited a simple argumentation level was significantly higher in the low-prior-knowledge group than that in the high-prior-knowledge group for all tasks (Table II).

For the argument accuracy, the group (2) \times task type (2) mixed ANOVA was conducted. Results demonstrated that the accuracy of the high-prior-knowledge group ($M = 0.97$) was significantly higher than that of the low-prior-knowledge group ($M = 0.90$), $F(1, 94) = 5.58$, $p = 0.020$. However, we did not observe a main effect of task type, $F(1, 94) = 0.66$, $p = 0.419$, or of interaction

effects, $F(1, 94) = 0.01$, $p = 0.921$ (Table II). These results indicated that the high-prior-knowledge group exhibited more complex argumentation levels and had higher accuracy than the low-prior-knowledge group, although the effect of task type and interaction were nonsignificant.

B. Eye-movement performance

Eye-movement data with a valid ratio below 40% or an apparent drift were excluded for all tasks. Therefore, valid data were collected from 92 (high-prior-knowledge and low-prior-knowledge, $n = 47$ and 45), 94 (high-prior-knowledge and low-prior-knowledge, $n = 48$ and 46), 94 (high-prior-knowledge and low-prior-knowledge, $n = 48$ and 46), and 94 (high-prior-knowledge and low-prior-knowledge, $n = 48$ and 46) participants for each task, respectively, for inclusion in the analyses. Before conducting the main analysis, we compared the TFD of the two groups. A group (2) \times task type (4) mixed ANOVA on TFD revealed no interaction effect between the group and task type, $F(3, 267) = 0.85$, $p = 0.407$. However, the high-prior-knowledge group ($M = 92.73$) had a significantly longer TFD than did the low-prior-knowledge group ($M = 70.38$), $F(1, 89) = 7.44$, $p = 0.008$. Furthermore, the main effect was also significant for task type, $F(3, 267) = 18.06$, $p < 0.001$; see the left half of Table III. We adopted Scheffe's method to compare the TFD of the argumentation evaluation tasks with the argumentation generation task. The results revealed an F value of 52.60, which was significantly higher than the critical value ($F' = 11.34$), demonstrating that the TFD of the argumentation generation task was significantly longer than that of the argumentation evaluation tasks.

1. Transition numbers between the representations

To verify hypothesis 2-1, a group (2) \times task type (4) mixed ANOVA was performed on transition numbers to examine whether individuals with more prior knowledge would exhibit more transitions between representations, especially during more difficult tasks. In other words, we investigated the presence of a group-task type interaction. Results revealed no interaction effect, $F(3, 282) = 0.83$, $p = 0.478$. However, the high-prior-knowledge group ($M = 35.75$) had a significantly more transition numbers than did the low-

TABLE II. Numbers of participants in each argumentation level and their accuracy by groups.

	Level		Accuracy M (SD)
	Simple level	Complex level	
Argumentation evaluation task			
Low-prior-knowledge group	23	24	0.89 (0.19)
High-prior-knowledge group	11	38	0.95 (0.14)
Argumentation generation task			
Low-prior-knowledge group	29	18	0.91 (0.28)
High-prior-knowledge group	12	37	0.98 (0.14)

TABLE III. Means and standard deviations of TFD and transition numbers for tasks and representations as a function of the group.

Task	TFD	Transition numbers
	<i>M</i> (SD)	<i>M</i> (SD)
Identification task		
Low-prior-knowledge group	56.54 (35.87)	17.60 (16.22)
High-prior-knowledge group	79.05 (54.72)	24.41 (22.24)
Correlation task		
Low-prior-knowledge group	52.74 (33.50)	20.28 (21.49)
High-prior-knowledge group	79.94 (38.96)	31.35 (22.18)
Prediction task		
Low-prior-knowledge group	74.63 (45.67)	21.02 (17.74)
High-prior-knowledge group	83.79 (42.50)	23.86 (19.14)
Argument generating task		
Low-prior-knowledge group	97.63 (60.77)	50.30 (40.56)
High-prior-knowledge group	128.14 (108.55)	63.37 (41.88)

prior-knowledge group ($M = 27.30$), $F(1, 94) = 6.24$, $p = 0.014$. Moreover, the main effect was also significant on task type, $F(3, 282) = 46.11$, $p < 0.001$; see the right half of Table III. We employed Scheffe’s method to compare the transition numbers of the argumentation evaluation tasks with the argumentation generation task. The results revealed an F value of 137.21, which was significantly higher than the critical value ($F' = 11.34$), indicating that the transition numbers of the argumentation generation task were significantly higher than for the argumentation evaluation tasks.

Furthermore, we suspected that the greater transitions of the generation task might be due to its significantly longer TFD. Thus, we divided transitions by TFD to reveal their numbers every second to exclude the alternative explanation, and found a similar result—it was significantly higher in the generation task than in the evaluation one.

2. Attention distributions for different AOIs

To verify hypothesis 2-2, a group (2) \times task type (3) \times AOI (4) mixed ANOVA on the proportion of fixation durations was performed to examine whether individuals with rich prior knowledge distributed more attention to the appropriate representations when evaluating arguments. We thus investigated the interaction between task types and AOIs, especially in the high-prior-knowledge group. Results revealed an interaction effect among the three variables, $F(6, 540) = 4.84$, $p < 0.001$. Therefore, a simple interaction effect analysis was performed to further understand these relationships. A significant interaction effect existed between task type and AOI in the low-prior-knowledge group, $F(6, 540) = 3.33$, $p = 0.003$. The high-prior-knowledge group also exhibited a significant interaction effect between task type and AOI, $F(6, 540) = 8.39$, $p < 0.001$. As illustrated in Table IV, the effect of all AOIs was observed in the high-prior-knowledge group. The high-prior-knowledge group distributed significantly more attention to the table and equation during the identification task and the correlation task than during the prediction task, $F_s(2, 720) < 9.82$, $p < 0.021$. The high-prior-knowledge group spent a significantly longer fixation duration on the figure during the prediction task than during the other tasks, $F(2, 720) = 10.04$, $p < 0.001$. By contrast, the low-prior-knowledge group’s attention distribution was only significantly higher for the equation during the prediction task compared with the identification task and the correlation task, $F(2, 720) = 6.92$, $p < 0.001$.

To verify hypothesis 2-3, a group (2) \times AOI (4) mixed ANOVA on the proportion of fixation durations was performed to examine whether individuals with high-prior-knowledge distributed more attention to the representation table to generate arguments. We thus assessed the interaction between group and AOI. Results showed no interaction effect between the two factors, $F(3, 264) = 1.27$, $p = 0.285$, but we determined that both groups had a

TABLE IV. Means and standard deviations of the TFD ratio for tasks and AOIs as a function of the group.

Task	Question stem	Table	Figure	Equation
	<i>M</i> (SD)	<i>M</i> (SD)	<i>M</i> (SD)	<i>M</i> (SD)
Identification task				
Low-prior-knowledge group	0.35 (0.14)	0.33 (0.17)	0.21 (0.15)	0.11 (0.11)
High-prior-knowledge group	0.31 (0.12)	0.33 (0.19)	0.16 (0.11)	0.20 (0.14)
Correlation task				
Low-prior-knowledge group	0.31 (0.12)	0.36 (0.17)	0.24 (0.16)	0.09 (0.09)
High-prior-knowledge group	0.22 (0.09)	0.38 (0.19)	0.20 (0.14)	0.020 (0.15)
Prediction task				
Low-prior-knowledge group	0.32 (0.12)	0.30 (0.20)	0.20 (0.15)	0.18 (0.14)
High-prior-knowledge group	0.32 (0.12)	0.27 (0.15)	0.27 (0.18)	0.14 (0.11)
Argument generating task				
Low-prior-knowledge group	0.14 (0.08)	0.51 (0.15)	0.23 (0.14)	0.12 (0.09)
High-prior-knowledge group	0.12 (0.07)	0.49 (0.16)	0.22 (0.12)	0.17 (0.11)

significantly longer fixation duration for the table compared with the other AOIs, $F(3,264) = 147.30$, $p < 0.001$. Besides, the high-prior-knowledge group allocated significantly more attention to the equation than the low-prior-knowledge group, $F(1, 88) = 5.46$, $p = 0.022$.

3. Reading sequence for both groups

Figure 2 presents the reading sequences for the high-prior-knowledge and low-prior-knowledge groups in the four tasks. The Z-value matrix is displayed in the Appendix A. We determined that both groups were inclined to read back and forth between the question stem and one of the representations in the argumentation evaluation tasks, whereas they demonstrated unidirectional reading

pathways from the question stem to other representations in the argumentation generation task. This finding revealed that both groups understood how to process various task types, including referencing the statements and representations frequently to evaluate whether the statements were true and distributing more attention to the representations to generate arguments.

In the argumentation evaluation tasks, the high-prior-knowledge group frequently related the information between the question stem and the equation and between the table and the figure, and their reading sequences were consistent among the three tasks, as illustrated in the left half of Fig. 2. For example, in the prediction task, the transfer probabilities for the question stem to the equation

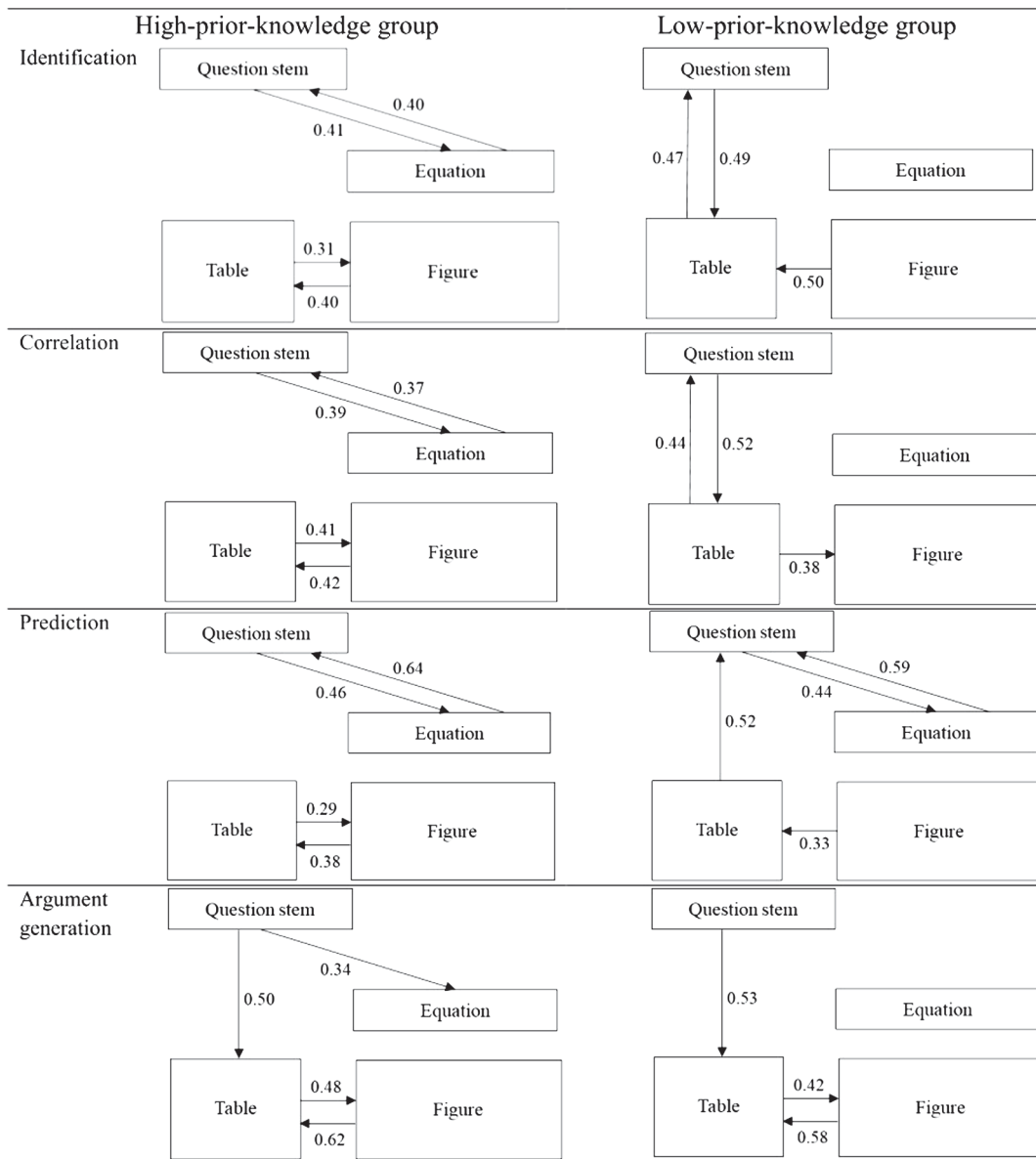


FIG. 2. Transition diagrams of the two groups in the four tasks. Note the arrows represent significant transition probabilities. The numbers beside the arrow indicate the transition probabilities.

($Z = 5.12$, $p < 0.001$) and for the equation to question stem ($Z = 12.41$, $p < 0.001$) were higher than for other AOIs. The transfer probabilities for the table to the figure ($Z = 2.91$, $p = 0.004$) and the figure to the table ($Z = 6.82$, $p < 0.001$) were higher than for other AOIs. The low-prior-knowledge group demonstrated similar reading sequences in the first two tasks, looking back and forth between the question stem and the table, as illustrated in the top right of Fig. 2. However, they referenced the question stem and the equation more in the prediction task. The transfer probabilities for the question stem to the equation ($Z = 3.44$, $p < 0.001$) and for the equation to question stem ($Z = 8.62$, $p < 0.001$) were higher than for other AOIs.

In the argumentation generation task, the two groups exhibited similar reading sequences. They seldom reread the information within the question stem. Instead, their reading sequences were often unidirectional from the question stem to the table, followed by a back and forth between the table and the figure. The high-prior-knowledge group also performed unidirectional behaviors from the question stem to the equation, as illustrated in the bottom left of Fig. 2. The transfer probability of the question stem to the equation was significantly greater than for other AOIs ($Z = 4.25$, $p < 0.001$).

IV. DISCUSSION AND CONCLUSIONS

We investigated argumentation performance and eye-movement behaviors when college students with high-prior and low-prior knowledge solved two types of multirepresentational tasks. Instead of relying solely on a verbal protocol, which can result in nonveridicality (see also Ref. [7]), this study used nonintrusive, real-time eye tracking, as a complementary approach to assess attention distribution among multiple representations when participants engaged in scientific argumentation tasks. For hypothesis 1-1 and 1-2, the results indicated that the number of participants who proposed complex arguments was greater in the high-prior-knowledge group than in the low-prior-knowledge group. Furthermore, the high-prior-knowledge group exhibited significantly higher argumentation accuracy than the low-prior-knowledge group did. However, no interaction effect or main effect of task type was found. For hypothesis 2-1, transitions between the representations were more frequent in the high-prior-knowledge group than that in the low-prior-knowledge group and in the argumentation generation task than that in the argumentation evaluation tasks. However, no interaction effect was observed. For hypothesis 2-2, we determined that both groups demonstrated an interaction effect between task types and AOIs in the argumentation evaluation tasks. For hypothesis 2-3, both groups had significantly more attention distribution for the table than the other AOIs during the argumentation generation task. Furthermore, we determined that the high-prior-knowledge group allocated significantly

more attention to the equation than the low-prior-knowledge group. For the exploratory research question, during the argumentation evaluation tasks, the high-prior-knowledge group was inclined to frequently transfer fixations between the question stem and the equation and between the table and the figure whereas the low-prior-knowledge group tended to only look back and forth between the question stem and the table. However, the low-prior-knowledge group exhibited a different reading sequence only for the prediction task. In the argumentation generation task, the two groups displayed similar reading sequences.

Hypothesis 1-1 and 1-2 are partially supported because the high-prior-knowledge group constructed higher quality and more accurate arguments than did the low-prior-knowledge group; however, the effects of task type were nonsignificant. This finding is in accord with those of Inglis and Alcock [7], von der Mühlen *et al.* [8], and Tsai *et al.* [27], who reported that individuals with more relevant knowledge could comprehend, integrate, and apply information more easily; thus, their argumentation performance was better. We assume that high-prior-knowledge students were more familiar with various scientific texts that contained relevant background knowledge and argument structure, which facilitated their argumentation performances.

Hypothesis 2-1 is partially supported because the high-prior-knowledge group demonstrated more transitions between representations than the low-prior-knowledge group did, and the generation task was also induced more transitions than the evaluation task, although no interaction effect between the two variables was observed. We found identical results after excluding the possible effect of TFD, that is, the number of transitions in every second. These findings indicated that the high-prior-knowledge students were more inclined to infer and integrate information between multiple representations compared with low-prior-knowledge students. It is consistent with studies [11,17,20] that have reported that high-prior-knowledge individuals were skilled at flexibly and fluidly moving across, connecting, or transforming representations of various forms.

Based on the findings of hypotheses 1-1, 1-2, and 2-1, we found there was no firm evidence to conclude that the perceived task difficulty of the generation task was higher than the evaluation ones. However, both groups did demonstrate more transitions and transitions per second, within the generation task. It indicated that individuals were more inclined to integrate information among multiple representations while generating scientific argumentations rather than evaluating them.

Hypotheses 2-2 and 2-3 are partially supported. A significant interaction effect between task type and AOI for the high-prior-knowledge group was observed. The high-prior-knowledge group distributed significantly more attention to the table and the equation in the identification task and the correlation task. Inconsistent with our prediction, we determined that the high-prior-knowledge

group had a significantly higher proportion of fixation durations for the figure during the prediction task. In contrast, the low-prior-knowledge group exhibited significantly higher fixation duration for the equation during the prediction task, which may have been because the equation in the prediction task was much more simple than in the other tasks, and thus the low-prior-knowledge group could infer the answer. This explanation was justified in our additional study. The additional study was conducted by collecting data on participants' perceived difficulty and willingness to use each representation according to questionnaire answers on a 7-point Likert scale. The participants were 10 individuals with low levels of prior knowledge involved in the formal study. The results revealed that the participants perceived higher difficulty in the equations ($M = 4.33$) than in the table ($M = 2.50$) and in the figures ($M = 3.10$). Participants also exhibited lower willingness to use equations ($M = 4.09$) than other representations (table and figures, $M = 5.10$ and 4.90). Moreover, we found differences in one representation across multiple tasks. The participants expressed that the equation in the prediction task ($M = 3.30$) was far easier than the equations in the other tasks, and they exhibited a higher tendency to use this equation ($M = 5.10$). Regarding hypothesis 2-3, both groups had a significantly longer fixation duration for the table than the other representations, indicating that all participants understood that the table contained the most information and was thus most suitable for generating arguments. Contrary to our expectations, we found that participants with high levels of prior knowledge did not devote any more attention to what we defined "appropriate representations" in all tasks. Instead, they tended to use multiple representations to complete the argumentation tasks and exhibited argumentation performance of higher quality and accuracy. These results indicated that individuals with an abundance of knowledge should possess higher representational flexibility in all types of tasks. The results also supported those of Klein *et al.* [11], Kohl and Finkelstein [17], and Kozma [20].

For the exploratory research question, the group with high prior knowledge performed reading sequences with consistent patterns in the argumentation evaluation tasks. They frequently moved between the question stem and the equation and between the table and the figure, indicating that they paid attention to multiple homogeneous representations more frequently than multiple heterogeneous representations. It is inconsistent with findings by Ott *et al.* [19], which may be because their participants were all novice university students. We determined that high-prior-knowledge university students may be more aware of the inter-representational coherence, which facilitated the translation processes across representations. The referencing behaviors between the representations text and equation might exert a constraining function [18], and the reference representation (i.e., text), which is more familiar for high-

prior-knowledge participants, constrained the interpretation of the abstract equation and thus improved their performance. As for the representation figure and table, although they are not typical analog representations, we suggested that they were similar code types because a point in the figure could also be identified from the data table. The referencing behaviors between the two representations may also have a constraining function [18] because the representation figure enables visualization of the trend and thus constrains meaning extraction from raw data (i.e., representation table), which fosters understanding. Another alternative explanation could exist—because the connected pairs were separated by smaller distances, the referencing behaviors may have been affected by the spatial contiguity principle [37]. On the other hand, the inspecting behaviors of the low-prior-knowledge group's reading sequence in the first two argumentation evaluation tasks were observed between the representation text and table, indicating that they spent more cognitive effort in understanding the task demands and searching for more information. Furthermore, the low-prior-knowledge group demonstrated different reading sequences in the prediction task, in which they referenced the question stem and the equation frequently. As revealed by our additional study, because the prediction task contained the simplest equation, the students with lower levels of prior knowledge were more willing to extract information from the equation.

Similar reading sequences were demonstrated by both groups in the argumentation generation task. They displayed one-way reading sequences from the question stem to the table and demonstrated referencing behaviors between the figure and the table. Because the task was open-ended, individuals tended to acquire more evidence for argumentations by inspecting the table, which contained abundant information. Furthermore, the representation figure provided a constraining range for understanding the complex raw data. The representation figure was beneficial for identifying the underlying patterns from the representation table and thus obtain a deeper understanding. These findings correspond to the constraining and constructing functions in the framework by Ainsworth [18]. These findings indicated that students with different prior knowledge had similar reading sequences when generating scientific argumentations. They generally determined the demands of the tasks and identified evidence from tables and figures. Furthermore, the group with high levels of prior knowledge also had one-way reading sequences from the question stem to the equation, whereas the group with low levels of prior knowledge did not inspect the equation, despite it being an easy linear equation.

Generally, the group with high levels of prior knowledge was able to select representations in various tasks with more flexibility. The data of the proportion of fixation durations demonstrated that the high-prior-knowledge group distributed attention to seemingly complicated equations and proposed arguments using more than one representation. This finding indicated that the high-prior-knowledge

learners were more able to obtain information from each representation and select the appropriate evidence to support their arguments. Furthermore, the high prior knowledge group’s frequent transitions between representations and the inspection of multiple homogenous representations may indicate an awareness of inter-representational coherence formation as well as an ability to integrate multiple representations in proposing complex arguments with higher accuracy or may bear a little relationship to spatial contiguity principle [37].

These findings have several pedagogical implications. First, we determined that the low-prior-knowledge group often engaged in unidirectional behaviors between homogeneous multiple representations (e.g., representation table and figure). Educators should provide low-prior-knowledge learners with homogeneous multiple representations and guide them in exploring the similarities between these representations, which would foster deeper learning and better argumentation performance. Second, this study determined that the low-prior-knowledge group paid limited attention to equations except the easiest one in the prediction task, revealing that lower-prior-knowledge learners used the equation less if the coefficient appeared complex. Therefore, learners should be reminded not to be intimidated and simplify the seemingly complex coefficients (e.g., make up a round number). Third, we determined that approximately 30%–40% of the participants provided only simple arguments, indicating that they may lack an understanding of scientific argumentation norms. Therefore, teachers should be supported in restructuring regular classroom activities and incorporating scientific argumentation into science learning.

This study had some limitations that should be addressed in the future. First, we investigated students’ argumentation

performances in various tasks with identical multirepresentational displays and their perceived task difficulty was not explicitly measured. Future studies should examine this using different combinations of representations and their perceived task difficulty to clarify the most used code types and the effects of tasks in physics learning. Second, we revealed that subtle differences within a certain representation (e.g., whether the coefficient of an equation was an integer or a decimal) might have affected the difficulty and suitability of the representation. Although the difficulty of representations may be difficult to control or independently manipulate, in future studies, experimental materials should be designed with more caution. Third, this study investigated *how* high- and low-prior-knowledge learners perform argumentations and *where* they pay attention to in a multiple-representational display but not *why* they select those representations. For example, the reason that the representation figure was attended frequently by the high-prior-knowledge group in the prediction task remains unclear. Future researchers could interview participants or request that they think aloud to further clarify the cognitive processes underlying individuals’ argumentation performance.

ACKNOWLEDGMENTS

This research was financially supported by the Ministry of Science and Technology (Project No. MOST 107-2511-H-003 -025) as well as the “Institute for Research Excellence in Learning Sciences” of National Taiwan Normal University (NTNU) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

APPENDIX A: POSSIBILITIES FOR TRANSITING FROM ONE AREA TO ANOTHER (Z-VALUE MATRIX)

The z-value matrix below presents the possibilities of transiting from the start area to target area in terms of the prior knowledge group.

Target area	Identification task			
	Question stem	Table	Figure	Equation
<i>Start area</i>				
<i>High-prior-knowledge group</i>				
Question stem		0.60	-4.12	3.46**
Table	1.18		2.33*	-3.12
Figure	-1.74	2.48*		-1.29
Equation	2.49*	-0.43	-2.17	
<i>Low-prior-knowledge group</i>				
Question stem		2.64**	-3.71	1.54
Table	3.56**		0.50	-5.05
Figure	-1.31	2.30*		-1.39
Equation	-0.11	-0.31	0.38	

Correlation task				
Target area	Question stem	Table	Figure	Equation
Start area				
<i>High-prior-knowledge group</i>				
Question stem		1.28	-5.10	4.00***
Table	-0.29		5.15***	-4.09
Figure	-4.59	3.25**		-0.04
Equation	3.42**	-2.56	-0.50	
<i>Low-prior-knowledge group</i>				
Question stem		2.94**	-4.75	-0.90
Table	3.18**		2.89**	-4.62
Figure	-1.42	1.58		-0.52
Equation	0.73	0.86	-1.35	
Prediction task				
Target area	Question stem	Table	Figure	Equation
Start area				
<i>High-prior-knowledge group</i>				
Question stem		-0.58	-3.35	5.12***
Table	-1.82		2.91**	-4.06
Figure	-0.99	6.82***		-6.03
Equation	12.41***	-6.15	-5.40	
<i>Low-prior-knowledge group</i>				
Question stem		1.47	-6.67	3.44**
Table	5.17***		1.84	-6.38
Figure	-2.59	2.08*		0.29
Equation	8.62***	-7.84	-1.39	
Argumentation generation task				
Target area	Question stem	Table	Figure	Equation
Start area				
<i>High-prior-knowledge group</i>				
Question stem		4.28***	-8.40	4.25***
Table	-1.60		7.04***	-4.59
Figure	-9.89	8.96***		-4.45
Equation	0.38	1.20	-0.70	
<i>Low-prior-knowledge group</i>				
Question stem		5.57***	-7.62	1.53
Table	0.04		3.17**	-2.50
Figure	-8.22	7.27***		-1.40
Equation	-1.32	0.67	0.09	

Note the rows indicate the starting areas and the columns indicate the target areas.

* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

APPENDIX B: THE TASKS USED IN THIS STUDY

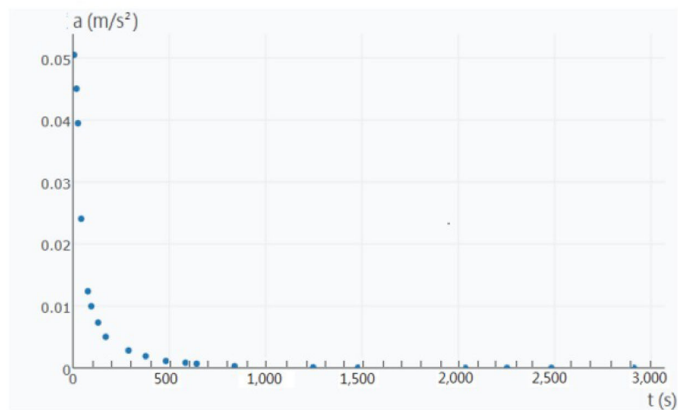
Four tasks was adopted in this study, including the identification task, the correlation task, the prediction task, and the argumentation generation task. Please see these materials below.

Identification task

Question: Thor believed that the acceleration of the weather balloon was initially 0.05 m/s², followed by a gradual decrease to 0 m/s². Please indicate that whether this statement is true and support your answer with the data.

Variable	Height	Time	Speed	Acceleration
Abbreviation	d	t	v	a
Unit	m	s	m/s	m/s ²
Data of the weather balloon	82.00	5.20	4.15	0.05
	271.00	18.40	4.56	0.05
	313.00	26.20	4.78	0.04
	400.00	42.90	4.94	0.02
	575.00	77.60	4.98	0.01
	664.00	95.20	4.99	0.01
	846.00	130.70	5.03	0.01
	1031.00	170.20	4.95	0.01
	1611.00	288.40	4.93	0.00
	2018.00	377.60	4.85	0.00
	2445.00	482.50	4.68	0.00
	2894.00	584.40	4.63	0.00
	3126.00	641.80	4.58	0.00
	3859.00	838.70	4.38	0.00
	5510.00	1246.50	4.27	0.00
	6456.00	1478.80	4.24	0.00
	8684.00	2035.30	4.18	0.00
	9566.00	2252.10	4.17	0.00
	10547.00	2782.50	4.17	0.00
	12318.00	2895.90	4.18	0.00

$$a = f(t) = -0.0000001t + 0.0003 \quad \text{if } t > 1000$$

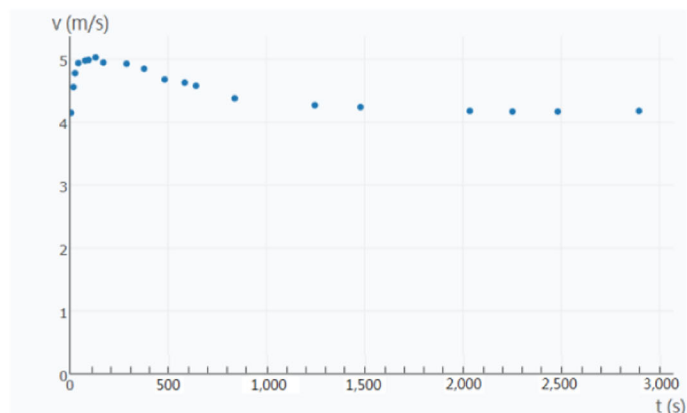


Correlation task

Question: Hulk proposed that the weather balloon was moving at a constant speed over time. Please indicate that whether this statement is true and support your answer with the data.

Variable	Height	Time	Speed	Acceleration
Abbreviation	d	t	v	a
Unit	m	s	m/s	m/s ²
Data of the weather balloon	82.00	5.20	4.15	0.05
	271.00	18.40	4.56	0.05
	313.00	26.20	4.78	0.04
	400.00	42.90	4.94	0.02
	575.00	77.60	4.98	0.01
	664.00	95.20	4.99	0.01
	846.00	130.70	5.03	0.01
	1031.00	170.20	4.95	0.01
	1611.00	288.40	4.93	0.00
	2018.00	377.60	4.85	0.00
	2445.00	482.50	4.68	0.00
	2894.00	584.40	4.63	0.00
	3126.00	641.80	4.58	0.00
	3859.00	838.70	4.38	0.00
	5510.00	1246.50	4.27	0.00
	6456.00	1478.80	4.24	0.00
	8684.00	2035.30	4.18	0.00
	9566.00	2252.10	4.17	0.00
	10547.00	2782.50	4.17	0.00
	12318.00	2895.90	4.18	0.00

$$v = f(t) = 0.0001t + 4 \quad \text{if } t > 1000$$



Prediction task

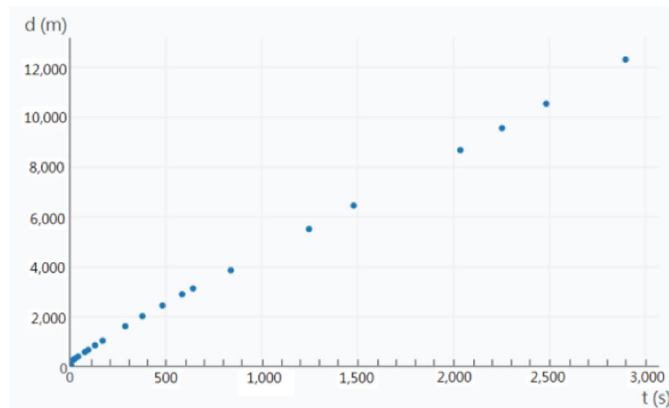
Please see Fig. 1. for this task.

Argumentation generation task

Question: Please observe this phenomenon, tell me about your claims in two to four sentences, and provide evidence for your claims.

Variable	Height	Time	Speed	Acceleration
Abbreviation	d	t	v	a
Unit	m	s	m/s	m*s*s
Data of the weather balloon	82.00	5.20	4.15	0.05
	271.00	18.40	4.56	0.05
	313.00	26.20	4.78	0.04
	400.00	42.90	4.94	0.02
	575.00	77.60	4.98	0.01
	664.00	95.20	4.99	0.01
	846.00	130.70	5.03	0.01
	1031.00	170.20	4.95	0.01
	1611.00	288.40	4.93	0.00
	2018.00	377.60	4.85	0.00
	2445.00	482.50	4.68	0.00
	2894.00	584.40	4.63	0.00
	3126.00	641.80	4.58	0.00
	3859.00	838.70	4.38	0.00
	5510.00	1246.50	4.27	0.00
	6456.00	1478.80	4.24	0.00
	8684.00	2035.30	4.18	0.00
	9566.00	2252.10	4.17	0.00
	10547.00	2782.50	4.17	0.00
	12318.00	2895.90	4.18	0.00

$$d = f(t) = 4t + 50 \quad \text{if } t > 1000$$



- [1] S. Fuller, *Science* (Open University Press, United Kingdom, 1997).
- [2] S. Erduran, S. Simon, and J. Osborne, TAPing into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse, *Sci. Educ.* **88**, 915 (2004).
- [3] R. Driver, P. Newton, and J. Osborne, Establishing the norms of scientific argumentation in classrooms, *Sci. Educ.* **84**, 287 (2000).
- [4] A. Kilinc, U. Demiral, and T. Kartal, Resistance to dialogic discourse in SSI teaching: The effects of an argumentation-based workshop, teaching practicum, and induction on a preservice science teacher, *J. Res. Sci. Teach.* **54**, 764 (2017).
- [5] *OECD, PISA 2012 results: What Students Know and Can Do—Student Performance in Mathematics, Reading and Science (Vol. 1)* (PISA, OECD Publishing, Paris, 2014).
- [6] J. Grooms, V. Sampson, and P. Enderle, How concept familiarity and experience with scientific argumentation are related to the way groups participate in an episode of argumentation, *J. Res. Sci. Teach.* **55**, 1264 (2018).
- [7] M. Inglis and L. Alcock, Expert and novice approaches to reading mathematical proofs, *J. Res. Math. Educ.* **43**, 358 (2012).
- [8] S. von der Mühlen, T. Richter, S. Schmid, and K. Berthold, How to improve argumentation comprehension in university students: Experimental test of a training approach, *Instr. Sci.* **47**, 215 (2019).
- [9] R. Duit, H. Niedderer, and H. Schecker, Teaching physics, in *Handbook of Research on Science Education*, edited by N. Lederman and S. Abell (Routledge/Taylor & Francis, London, 2014), pp. 434–456.
- [10] P. Klein, J. Viiri, and J. Kuhn, Visual cues improve students' understanding of divergence and curl: Evidence from eye movements during reading and problem solving, *Phys. Rev. Phys. Educ. Res.* **15**, 010126 (2019).
- [11] P. Klein, J. Viiri, S. Mozaffari, A. Dengel, and J. Kuhn, Instruction-based clinical eye-tracking study on the visual interpretation of divergence: How do students look at vector field plots?, *Phys. Rev. Phys. Educ. Res.* **14**, 010116 (2018).
- [12] W. Schnotz, U. Ludewig, M. Ullrich, H. Horz, N. McElvany, and J. Baumert, Strategy shifts during learning from texts and pictures, *Jpn. J. Educ. Psychol.* **106**, 974 (2014).
- [13] A. M. Madsen, A. M. Larson, L. C. Loschky, and N. S. Rebello, Differences in visual attention between those who

- correctly and incorrectly answer physics problems, *Phys. Rev. Phys. Educ. Res.* **8**, 010122 (2012).
- [14] B. Cavlazoglu and C. Stuessy, Examining science teachers' argumentation in a teacher workshop on earthquake engineering, *J. Sci. Educ. Technol.* **27**, 348 (2018).
- [15] Y. C. Jian, Fourth graders' cognitive processes and learning strategies for reading illustrated biology texts: Eye movement measurements, *Read. Res. Q.* **51**, 93 (2016), <https://ila.onlinelibrary.wiley.com/doi/full/10.1002/rrq.125>.
- [16] M. Stieff, M. Hegarty, and G. Deslongchamps, Identifying representational competence with multirepresentational displays, *Cognit. Instr.* **29**, 123 (2011).
- [17] P. B. Kohl and N. D. Finkelstein, Patterns of multiple representation use by experts and novices during physics problem solving, *Phys. Rev. Phys. Educ. Res.* **4**, 010111 (2008).
- [18] S. Ainsworth, The Multiple representational principle in multimedia learning, in *The Cambridge Handbook of Multimedia Learning*, edited by R. Mayer (Cambridge University Press, Cambridge, England, 2014), pp. 464–486.
- [19] N. Ott, R. Brünken, M. Vogel, and S. Malone, Multiple symbolic representations: The combination of formula and text supports problem solving in the mathematical field of propositional logic, *Learn. Instr.* **58**, 88 (2018).
- [20] R. Kozma, The material features of multiple representations and their cognitive and social affordances for science understanding, *Learn. Instr.* **13**, 205 (2003).
- [21] J. Meyer, D. Shinar, and D. Leiser, Multiple factors that determine performance with tables and graphs, *Human Factors* **39**, 268 (1997).
- [22] F. R. Curcio, Comprehension of mathematical relationships expressed in graphs, *J. Res. Math. Educ.* **18**, 382 (1987).
- [23] N. Friel, F. R. Curcio, and G. W. Bright, Making sense of graphs: Critical factors influencing comprehension and instructional implications, *J. Res. Math. Educ.* **32**, 124 (2001).
- [24] S. Kriz and M. Hegarty, Top-down and bottom-up influences on learning from animations, *Int. J. Hum-Comput. St.* **65**, 911 (2007).
- [25] R. A. Duschl and J. Osborne, Supporting and promoting argumentation discourse in science education, *Stud. Sci. Educ.* **38**, 39 (2002).
- [26] Y. R. Lin, The influences of contextualized media on students' science attitudes, knowledge, and argumentation learning through online gam-based activities, *J. Comput. Assist. Learn.* **34**, 884 (2018).
- [27] P. Y. Tsai, T. T. Yang, H. C. She, and S. C. Chen, Leveraging college students' scientific evidence-based reasoning performance with eye-tracking-supported metacognition, *J. Sci. Educ. Technol.* **28**, 613 (2019).
- [28] S. Toulmin, *The Uses of Argument* (Cambridge University Press, Cambridge, England, 1958).
- [29] F. Y. Yang, Examining the reasoning of conflicting science information from the information processing perspective—an eye movement analysis, *J. Res. Sci. Teach.* **54**, 1347 (2017).
- [30] G. Beattie, M. Marselle, L. McGuire, and D. Litchfield, Staying over-optimistic about the future: Uncovering attentional biases to climate change messages, *Semiotica* **2017**, 21 (2017).
- [31] M. Haigh, J. S. Wood, and A. J. Stewart, Slippery slope arguments imply opposition to change, *Mem. Cogn.* **44**, 819 (2016).
- [32] M. H. Yen and Y. T. Wu, The role of university students' informal reasoning ability and disposition in their engagement and outcomes of online reading regarding a controversial issue: An eye tracking study, *Comput. Hum. Behav.* **75**, 14 (2017).
- [33] B. W. Wojdowski and H. Bang, Distraction effects of contextual advertising on online news processing: An eye-tracking study, *Behav. Inf. Tech.* **35**, 654 (2016).
- [34] D. Wu, H. Huang, N. Liu, and D. Miao, Information processing under high and low distractions using eye tracking, *Cogn. Process.* **20**, 11 (2019).
- [35] R. J. Beicher, Testing student interpretation of kinematics graphs, *Phys. Teach.* **62**, 750 (1994).
- [36] T. Erickson, Stealing from physics: Modeling with mathematical functions in data-rich contexts, *Teach. Math. Appl.* **25**, 23 (2006).
- [37] R. Mayer, Spatial contiguity principle, in *Multimedia Learning*, edited by R. Mayer (Cambridge University Press, Cambridge, England, 2009), pp. 135–152.