DOI:10.6129/CJP.202203_64(1).0003
中華心理學刊 民 111，64 卷，1 期，57-76
Chinese Journal of Psychology 2022, Vol.64, No.1, 57-76

Conveying Spatial and Kinematic Representations in Text Reading via Words and Pictures: An Eye-movement Analysis

Yu-Cin Jian and Chao-Jung Wu
Department of Educational Psychology and Counseling, National Taiwan Normal University

This study investigated how words and pictures convey spatial and kinematic representations in a reading condition and the relations between these two representations. Participants read a text (text group) describing the spatial configuration of a flushing cistern or viewed a diagram (diagram group) and completed a spatial test. Then, the two groups read the same text describing the kinematic machine operation and revised their first test answers. The participants' eye movements were recorded while reading the kinematic information text. According to the results, the text group's drawings showed the component relations more precisely, indicating that words helped readers capture detailed continuous relations, but the diagram group's drawings better approximated the original picture, indicating the advantage of depicting overall relations. Analysis of the drawing tests indicated that kinematic representations help readers revise inner spatial representations mainly of continuous relations of components rather than part-whole relations. Moreover, eye movement data indicated that the groups differed in both global and local eye movements, and that different internal spatial representations of the spatial and diagram groups influenced the following kinematic information text reading. The text group had significantly shorter total fixation duration and second-pass reading time on the kinematic text, fewer complicated concept sentences (i.e., outlet processes), and fewer target words for components than the diagram group. In summary, this study confirmed that pictures and words served fundamentally different functions in conveying spatial configurations and clarified the interaction of spatial and kinematic representations from reading to mental model construction.

Keywords: *eye movements, kinematic representation, reading, spatial configuration*

Machines, such as flushing cisterns, bicycle pumps, photocopiers, and washing machines, are highly relevant in daily lives. People often learn the functioning principles of these machines by reading instructional manuals, textbooks, or science texts that often comprise words and pictures. Text and pictures have different functions in cognitive processes (Schnotz & Bannert, 2003; Schnotz et al., 2014). The text usually serves as a conceptual guide for initial comprehension (Schnotz & Wagner, 2018), while pictures are used as a mental scaffold to facilitate mental model construction (Eitel et al., 2013; McCrudden et al., 2011). However, this study

Received: 2021/1/22; First Revision: 2021/4/20; Second Revision: 2021/7/4; Accepted: 2021/7/6:
Corresponding Author: Chao-Jung Wu(cjwu@ntnu.edu.tw) Department of Educational Psychology and Counseling, National Taiwan Normal University, No. 162, Heping East Road Section 1, Taipei 106, Taiwan, R.O.C.

Acknowledgement: This work was financially supported from the MOST108-2410-H-003-046-MY2 for Professor Chao-Jung Wu, both of the MOST109-2636-H-003-003- and MOST110-2636-H-003-003- under Young Scholar Fellowship Program by Ministry of Science and Technology (MOST) for Associate Professor Yu-Cin Jian, and the "Institute for Research Excellence in Learning Sciences" and "Higher Education Deep Cultivation Project" of National Taiwan Normal University (NTNU), sponsored by the Ministry of Education, Taiwan.

focused on the role of text and pictures in constructing a coherent mental model for comprehending machine operations while reading a machine article.

Constructing a kinematic representation of machine operation is an essential cognitive process for comprehending instructional manuals or texts for machines. Kinematic representation involves the formation of a coherent mental model in which physical activity changes sequentially with time (Ainsworth & VanLabeke, 2004). Machines have two characteristics: *spatial configuration* and *kinematic information* (Hegarty, 1992; Heiser & Tversky, 2006; Kriz & Hegarty, 2007). A reader must not only construct an internal kinematic representation (the functional organization) of machine operation, that is, imagining how the machine works, but also capture the spatial configuration (the structural organization) of the machine's components (Ainsworth & VanLabeke, 2004; Hegarty et al., 2003; Heiser & Tversky, 2006). Previous research have focused on kinematic representations (Jian & Wu, 2016; Jian et al., 2014; Kriz & Hegarty, 2007), but less attention has been paid to the spatial configuration and the interaction between the spatial and kinematic representations constructed by readers. Therefore, this study investigated how words and pictures compare in conveying the spatial configuration of a machine when reading, as well as the relations between the spatial and kinematic representations.

Text and the Picture Comprehension Model

Schnotz et al. constructed a *model of text and picture comprehension* (Schnotz & Bannert, 2003) and conducted a series of empirical studies to confirm the routes by which text and pictures are processed (Schnotz & Wagner, 2018; Schnotz et al., 2014). The *model* describes how the representations of text and pictures connect and integrate with each other. Reading to gain knowledge during illustrated text reading has two different cognitive processes: text and pictures serve as two external representations of visual inputs that respond to descriptive and depictive representations. According to this model, text comprehension initiates the construction of a representation of the text surface structure, generates a propositional representation of the semantic content, and finally forms a mental model of the text. Specifically, this text comprehension process is based on the activation of schemata, which involves the interaction of top-down and bottom-up processes. Cognitive schemata select and organize information relevant to the reading task, triggered by the activation of a top-down process. The selected information is first organized into a *text surface structure*, a coherent *propositional representation* is then formed through a semantic process, and finally a *mental model* is constructed. Propositional representations and mental models interact continuously, guided by the schemata. The interaction and switching of these descriptive representations at different levels rely on the cognitive process of symbol-structure analysis.

In picture comprehension, Schnotz and Wagner (2018) indicated that an individual first perceives external pictures, then creates a visual representation of the pictures through perceptual processing that organizes graphic entities in the visuo-spatial sketchpad in working memory, as per Gestalt Laws (Ohlsson, 1984), and finally constructs a propositional representation and mental model through semantic processing. The interactive switching of depictive representations at different levels depends on structural mapping processes. During this process, the graphical entities are mapped on mental entities, preserving the structural characteristics of the intended images. Spatial relations are mapped on semantic relations, integrating conceptually with an individual's prior knowledge.

Words or Pictures in Conveying Spatial Configurations

Some scholars (Larkin & Simon, 1987; Tversky, 2001) have suggested that diagrams are more efficient than verbal descriptions for conveying spatial configurations. Pictures reduce the need to search for multiple information elements related to a single idea, as this information is grouped in visualizations, allowing readers to perceive the information visually (Goldstone & Son, 2005; Schmidt-Weigand & Scheiter, 2011). By contrast, text has no natural mapping of diagrams for the elements and configuration of a system (Heiser & Tversky, 2006). However, some scholars have suggested that verbal descriptions are also effective in conveying spatial structure, as long as the text provides sufficient information (Medin et al., 2005; Taylor & Tversky, 1992).

To distinguish the differential processes in constructing spatial representation by viewing diagrams versus reading propositions, Medin et al. (2005) used a simple map of the relative locations of six cities as an example to illustrate. The spatial information of the location relations of any city can be easily constructed using the analogical representation of the map. However, many propositions are required to describe the spatial information of maps. One way needs 30 propositions (P_2^6) to describe the location relations of any two cities in the map. For example, city A is 30 miles south of city B, city B is 30 miles north of A, city A is 10 miles west of city C, and city C is 10 miles east of city A. There is another method that requires only five propositions to convey the spatial relations of the map. City A was 30 miles south of city B, city B 10 miles west of city C, 10 miles north of city D, city D is 20 miles south of city E, and city E is 10 miles west of city F. The two styles of propositional representations indicate the trade-off between memory loading and computation complexity (Greeno & Simon, 1974). Using 30 propositions requires high memory loading but low computational complexity. Readers do not need to perform spatial reasoning because the relations of any two cities are provided in the 30 propositions. However, the memory loading is very high. By contrast, using five propositions requires low memory loading but high computational complexity. Although readers need not memorize many spatial propositions, they need to infer the relative relations of any other two cities if the relations have not been provided directly (e.g., the relationship between city A and city D). This example indicates that constructing spatial representations from reading propositional representations is a complex process, but viewing analog representations is relatively simple.

Schmidt-Weigand and Scheiter (2011) provided evidence supporting the effectiveness of words in conveying spatial information. They randomly assigned undergraduate participants to one of four reading conditions to gain meteorology knowledge: high spatial information text, low spatial information text, high spatial information text plus animation, and low spatial information text plus animation. The reading material discussed the steps of the lightning formation. The high spatial text contained phrases of direction alteration (e.g., "The charge within the cloud is caused by collisions among the rising and falling water droplets"). The low spatial text removed the spatial phrases from the text (e.g., by changing "rising and falling air currents" into "air currents moving in opposite directions"). After reading, participants performed a retention test (measured degree of memorization), a transfer test (measured concept learning), and a drawing test (measured the learner's visual-spatial representation of the reading situations). The results showed that there was no multimedia effect for high spatial text for retention and visual tests; however, there was a strong multimedia effect for the low spatial text. This suggests that participants were able to construct spatial representations from reading pure text describing spatial information using many sentences, because the group that read high spatial information plus animation did not perform better than those who read pure text with high spatial information.

Heiser and Tversky (2006) also showed that adult readers were capable of constructing spatial representations by reading a structural description of a car brake. In Experiment 2, undergraduate students were instructed to sketch a diagram of what they believed the textual description of the mechanical system (car brake, bicycle pump, and pulley system) was trying to convey. The description included both structural descriptions, containing details of parts and their spatial relations, and functional descriptions, containing the actions and consequences of a dynamic event. The results showed that learners depicted many arrows indicating moving directions in their diagrams when they read the car brake's functional description; however, the learners depicted many correct components and labels indicating that they captured good spatial information when they read the car brake's structural description.

Moreover, from the viewpoint of linguistics, some writing systems, such as Chinese (used in this study), many words have high semantic transparency. That is, the meaning of a word is easily decoded from its constituent characters' meanings (Libben et al., 2003), which might convey spatial information. For example, 連接桿 "connecting rods" conveys the continuous relations of spatial configuration in a machine. Readers can infer that there must be two components in a machine connected to the rod: one at the starting point and the other at the endpoint of the rod. Similarly, 上圓盤 "upper disk" and 下圓盤 "lower disk" indicate relative positions in a machine.

In summary, the above studies have shown that words and diagrams may convey spatial configurations in different ways. Readers are capable of constructing some degree of spatial representation by reading word descriptions or viewing diagrams. However, few studies have evaluated the quality of the spatial representations that readers constructed or what they could and could not construct while reading textual spatial descriptions. For example, Heiser and Tversky (2006) only coded the numbers of arrows and lines drawn on the pictures, including the placement (inside or outside the diagram) and function (labeling, sequence, and motion), in order to address their research questions. If a reader drew an arrow in his or her picture, but the start and end components were incorrect, it indicated that the reader captured some concept of the kinematic information in the textual description, but their internal spatial representations of detailed connective relations of components were not entirely correct. Therefore, this study further refined the scoring criterion of spatial representation to determine whether a reader captured the global (the similarity of the whole entity) and partial (part-whole relations, relative positions of detailed components) spatial configurations while reading the textual description. Thus, we can determine whether readers can and cannot construct detailed spatial representations by reading words and viewing diagrams.

Kinematic Representations in Mechanical Operations

In the past two decades, several studies have used eye-tracing technology to investigate how learners construct kinematic representations from reading science articles (Boucheix & Lowe, 2010; Hegarty & Just, 1993; Jian & Wu, 2016) and science diagrams (Jian et al., 2014). Hegarty and Just (1993) sought to determine which information in a text and diagram is integrated to construct a kinematic representation of the operation of a pulley system. They asked undergraduate participants to read an illustrated text describing the configuration and kinematics of the pulley systems, and the participants' eye movements were recorded. The eye movement data showed that most readers first read the text increments describing the pulley system's operation and often reread sections about particular components or groups of components. They then looked at the diagram to integrate text and diagram information. When viewing the pulley's diagram, their inspections served to elaborate the configuration and kinematic relations described in the text. Overall, Hegary and Joust found that readers combined local representations of the components and global representations to construct their mental model of machine operation.

Jian et al. (2014) investigated how readers construct kinematic representations by reading diagrams with numbered arrows. This study used an eye-tracker and asked undergraduate participants to read a two-stage diagram depicting a flushing cistern with or without numbered arrows and answer questions. The readers in the arrow groups had better test scores for the kinematic concept measurements than the non-arrow groups. In addition, this study adopted a series of sequential analysis matrix calculations (Bakeman & Gottman, 1997; Jian, 2016; Jian & Ko, 2017; Tsai et al., 2019; Wu et al., 2021; Wu & Liu, 2021) of eye-fixation sequences, indicating the different processing strategies adopted by the arrow and non-arrow groups. The arrow group followed the numbered arrows to construct a kinematic representation of the flushing cistern operation, while the non-arrow group compared the different statuses of the two-stage (outlet and inlet) diagrams to infer how the flushing cistern operated. In another study, Jian and Wu (2016) investigated whether kinematic information conveyed via diagrams differed from that communicated by words. Undergraduate participants viewed diagrams (one version with numbered arrows and another without) depicting how a flush system operates, and then read an illustrated text describing the machine operation. The results showed that the arrow group performed better on the test measuring the concept of kinematic operation than the non-arrow group, and that the arrow group spent less time reading the diagram and text—which conveyed the concept in a much simpler (inlet-process) manner—than the non-arrow group; however, both groups spent more time reading complicated (outlet-process) concepts. These findings indicate that readers can construct a basic kinematic representation by reading numbered arrows, but constructing complicated kinematic representations requires reading textual descriptions of a machine's operation. That is, the kinematic information conveyed via diagrams may depend on that conveyed via words to some extent.

In summary, most studies on machine operation reading have measured how well readers form kinematic representations during reading (Hegarty, 1992; Hegarty & Just, 1993; Heiser & Tversky, 2006; Jian et al., 2014; Kriz & Hegarty, 2007; Mayer & Gallini, 1990). Few studies have evaluated the quality of the spatial representations readers constructed or what they can and cannot construct while reading spatial descriptions. Meanwhile, spatial configuration is related to kinematic operation in constructing a mental model for comprehending kinematic operation (Hegarty, 1992; Hegarty & Just, 1993). Thus, this study sought to address this gap by enriching the research on kinematic representation.

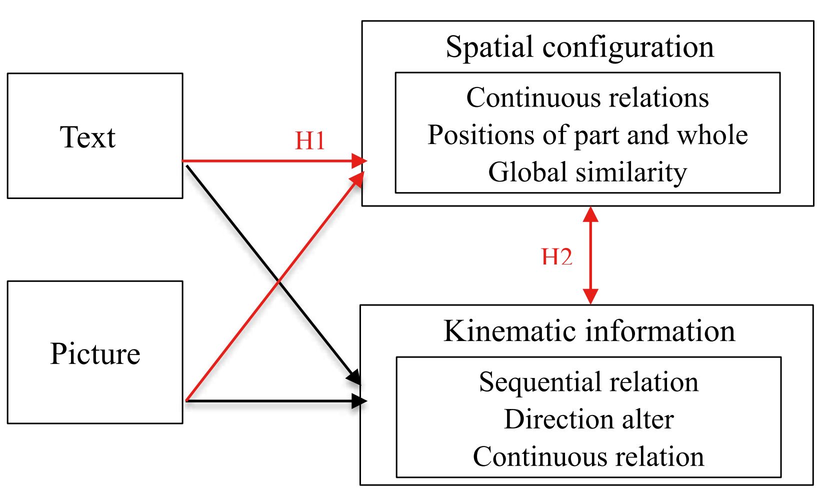
The Present Study

This study investigated two research questions (Figure 1). The first research question addressed the potential differences among different dimensions (e.g., continuous relations of components, locations of part-whole relations, and global similarity) of the spatial configurations that readers constructed when reading spatial information provided through text and pictures. The second research question addressed the interaction of spatial configuration and kinematic representation in reading, that is, whether the spatial configuration of a machine influences readers' construction of the kinematic representation, and conversely, whether readers' kinematic representation also influences spatial configuration formation.

To investigate the functions of "*words*" and "*pictures*" in describing the spatial configurations and kinematic information of a machine system separately, this study adopted a two-stage reading procedure (Figure 2), where spatial configuration information was read in the first stage and kinematic information in the second stage. The text group read a text describing the spatial

Figure 1

The two hypotheses in this study are indicated by the red lines. The two black line have been confirmed by Jian & Wu (2016) and Jian et al. (2014)



configuration of a flushing cistern, and the diagram group viewed a diagram of the flushing cistern. Both groups then read the same kinematic information text describing the operation process of a flushing cistern. While completing text reading or picture viewing in the first reading stage, participants answered yes-or-no questions on the detailed continuous relations of components in the machine system, and then drew a picture (scored on three dimensions: detailed continuous relations of components, locations of part-and-whole relations, and global similarity) by imagining the spatial configuration of the flushing cistern based on the information they read. In the second reading stage, the two groups read the same pure text describing the kinematic information of the flushing cistern's operation. The participants' eye movements were recorded to examine whether both groups (although with different spatial representations constructed from the first-stage reading task) had different reading processes when reading the kinematic description text. After reading the article, the participants were instructed to revise their answers and the picture they drew in the first stage.

Research Hypotheses

We formulated hypotheses for the research questions. For the first research question, because words and pictures play different roles in illustrated text reading (Schnotz & Bannert, 2003; Schnotz et al., 2014), we expected that this difference would manifest in the performance of the text and diagram groups on the spatial configuration tests. If words had an advantage in conveying detailed continuous relations (Medin et al., 2005; Schmidt-Weigand & Scheiter, 2011) and pictures had an advantage in conveying the overall relations of locations, shapes, and sizes (Larkin & Simon, 1987; Schnotz et al., 2014; Tversky, 2001), we expected that the two groups would show different concepts of spatial relations: the text group should outperform the diagram group on the spatial configuration test to measure the detailed continuous relations of the flushing cistern, but the diagram group should outperform the text group on the same test and score higher on the dimensions of global spatial relations and part-whole relations (Hypothesis 1).

For the second research question, we expected the spatial and diagram groups to have different eye movements during the kinematic text reading (Hypothesis 2a), because the two groups might construct different spatial representations from word reading or picture viewing in the first-stage reading; this would influence the subsequent reading of the kinematic text, describing how a flushing system works. Because a kinematic event requires at least two continuous components (e.g., the connecting rod pulls up the lower disk), if the text group better captures the components' continuous relationship in a flushing cistern in their mind, then in the following kinematic text reading, they would manipulate kinematic operation representations based on their spatial imagery (Hegarty & Tversky, 2006). Thus, we expected that there would be labor saving in the text group due to reading the kinematic text, relative to the diagram group.

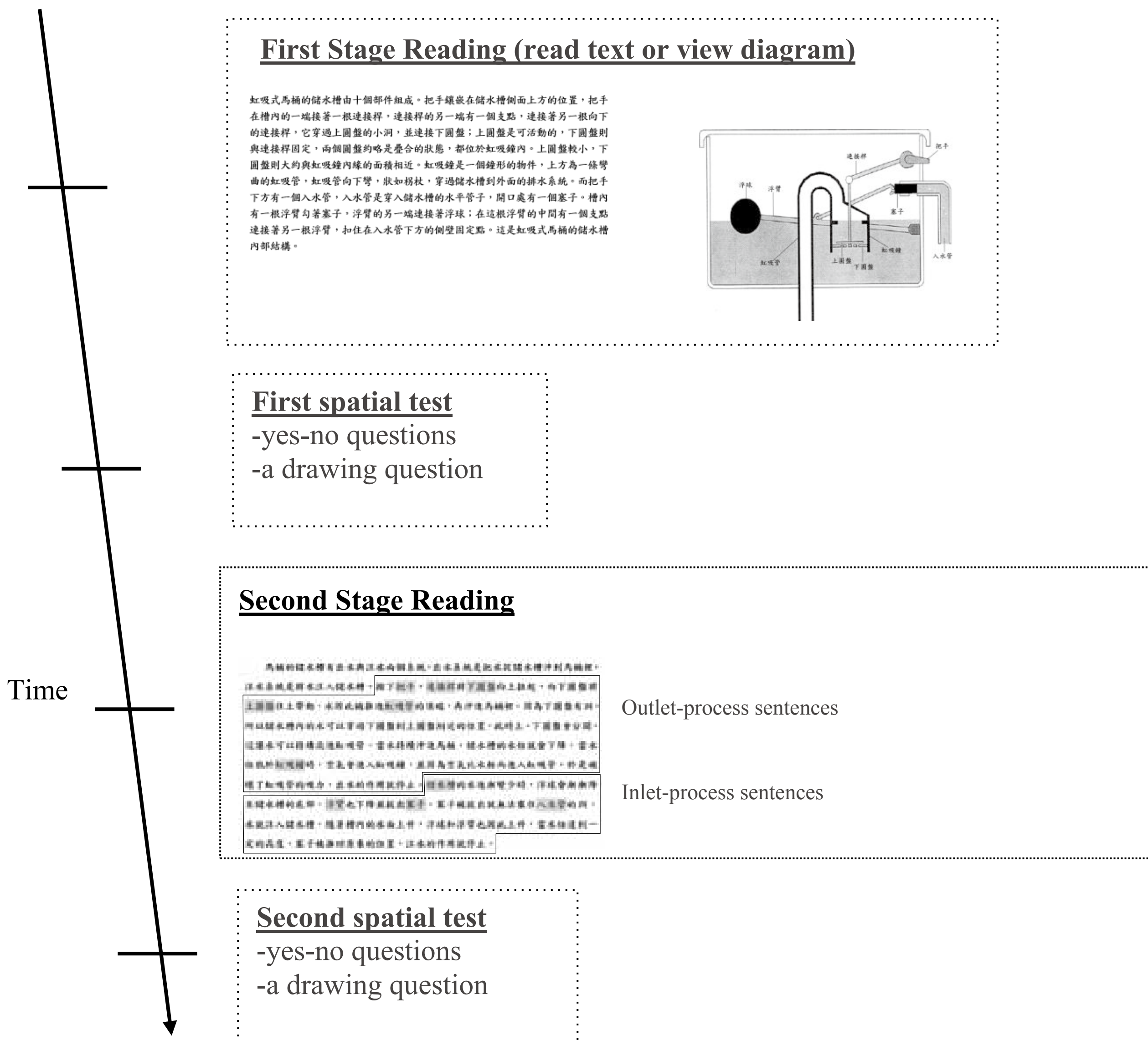
As for the expectation of an influence in the reverse direction of the kinematic representation of the spatial representation the participants constructed, we expected that the text group would score significantly better on the second revised spatial configuration test, especially on the dimensions of the detailed continuous components, than the diagram group (Hypothesis 2b). This is because a kinematic event requires at least two continuous components, as mentioned above, and this continuous relation is provided in the kinematic text.

Method

Participants

Forty-four undergraduate students (20 males and 22 females) from National Taiwan Normal University volunteered to participate for a small monetary reward. The mean age of the participants was 22.55 years (*SD* = 2.36), and they were students of education, management, the arts, or social science. We excluded students who majored in science or engineering, as they were expected to have minimal background knowledge in these areas. All participants were native speakers of the experimental materials in Chinese who had normal or corrected-to-normal vision.

Figure 2
The experimental procedures and materials (The black lines and gray ground in the second stage reading were areas of interests in eye-movements analyses, and didn't shown for the participants)



Materials and Design

There were two experimental materials (Figure 2) to be learned sequentially: the spatial configuration of a flushing cistern and a text describing how the flushing cistern works. The text group read a text describing the spatial configuration of the flushing cistern, including partial relations of 10 components (e.g., "*The handle links to the connecting rod*") and their locations relative to the flushing cistern (e.g., "*The siphon is downward and crosses over the flushing cistern to the drainage pipe*"), that is, their part-whole relations. The diagram group reads a diagram of the flushing cistern (Hegarty et al., 2003).

The kinematic information text was translated into

Chinese from the English narration (Kriz & Hegarty, 2007) by two researchers who majored in reading research. To confirm the readability of the article, 20 undergraduate students who did not participate in the formal experiment were recruited. This kinematic text was read by all the participants in the eye-tracking experiment. This text describes how the flushing cistern works, comprising 341 words that describe the steps of the outlet process (complicated concept) and the inlet process (simple concept) operations in its workings. The text comprised seven sentences. Sentence 1 briefly introduces the function of the outlet processes (how water flows out to the siphon pipe in the flushing cistern) and inlet processes (how water pours into the flushing cistern from the water pipe). Sentences 2–4 describe the principles of the outlet process and the steps involved. Sentences 5–7 describe the principles of the inlet process and the steps involved. All reading materials were individually displayed on a screen. There were no rolling bars to pull down.

Reading Tests

The reading tests measured the spatial configuration and kinematic information of the flushing cistern. Among them, the spatial configuration tests had two types of items measuring different levels of concepts: (1) yes-or-no items, which measured continuous relations of specific components (e.g., "*Is the connecting rod connected to the upper disk?*"; 10 items in total); and (2) drawing a picture of the spatial configuration of the flushing cistern, which measured the quality of the mental representation constructed by the reader while reading the text describing the machine's spatial configuration. Three scoring criteria were used for the images. The first was the continuous relations of components, measuring the detailed connective relations of components with each other. The second was the location of the components in the flushing cistern, measuring the relative relations of parts with the entire spatial configuration. The third is global similarity, which measures the similarity between the picture and the actual diagram of the flushing cistern. The scoring criteria included the components' locations, proportions, shapes, and sizes. We conducted a pilot study with 13 undergraduates to ensure that all the reading materials were fluent and readable, and that the pictures participants drew could be scored according to the three criteria. These students did not participate in the eye-tracking experiment.

Apparatus

The participants' eye movements were recorded using an Eyelink 1000 at a sampling rate of 1000 Hz. This system is accurate to a visual angle of 0.5 °. A chin bar was used to minimize head movements. Binocular and eye movements were recorded from the right eye only. The stimuli were presented on a 24-inch LCD monitor with a resolution of 1920 × 1200 pixels, and the entire text was visible on the screen; no page scrolling was required. The size of each Chinese character in the text displayed was 42 × 42 pixels; each character subtended a visual angle of 0.84 º at a distance of approximately 65 cm during the reading tasks.

Procedure

The participants were tested individually during the experiment. Before the formal experiment, all participants were instructed to read three expository texts (each 300–500 words in length, near that of the experimental material text) to ensure that the participants in the spatial and diagram groups had similar eye movement patterns when reading expository texts. The formal experiment involved a two-stage reading procedure. The first reading stage involved reading the spatial configuration information of the flushing cistern, and the second stage involved reading a text describing the kinematic information of the workings of the flushing cistern. Participants were randomly assigned to one of two groups based on reading conditions: the text group, which read a pure text describing the spatial configuration of the flushing cistern, or the diagram group, which viewed a diagram of the flushing cistern. Participants were instructed to read the learning materials in order to produce an image of the spatial configuration of the flushing cistern, and reading tests were completed. Reading time was not limited. Before the start of the experiment, a 12-point calibration and validation procedure was conducted, where the participants read a practice text first and answered several reading questions. Subsequently, the formal experiment was initiated.

After participants read the experimental materials, they completed the yes-or-no spatial configuration items and were instructed to draw a picture by imagining the spatial configuration described in the reading materials. The text group was allowed to draw the picture and read the texts at the same time, but the diagram group performed the drawing task without having the diagram present. Participants were instructed that the scoring criteria involved drawing the spatial structure and connecting the relations of the components, rather than making the drawing detailed and attractive. The drawing time was limited to 13 min based on a pilot study, confirming that this was sufficient for readers to complete the drawing.

Before starting the second reading stage, participants performed the 12-point calibration and validation procedures again. The two groups then read the same text describing the kinematic information of how the flushing cistern works without time limitations. They were instructed to press a keyboard to terminate the display when they finished reading, and then revise their answers on the spatial configuration test and the picture they drew, again without time limitations. In total, the experiment took approximately 50–60 min.

The rationale for this experimental procedure was based on three considerations. First, the purpose of this study was to investigate whether readers were capable of constructing the spatial configuration of a machine system by reading a verbal description; therefore, we instructed the participants to read the text with the goal of imagining its spatial configuration. Otherwise, the readers might have read the text to get its gist rather than to form a spatial configuration, ultimately failing to respond to the research questions. Second, we did not tell the participants that they could read the text again in the following drawing test to avoid unserious reading in the first stage. Third, as the purpose of this study was to investigate whether readers were capable of constructing a spatial imagery representation by reading words, rather than to test their memory, we gave them the spatial configuration text that they had read earlier while they were drawing the spatial picture. If we had not provided the original text for the readers in the text group while they were drawing, their poor results may be due to limited working memory, in which case we would not be able to distinguish the effects of memory and ability. We designed it in this manner because if the original diagram was presented to the participants of the diagram group and they performed better than the other two groups on the drawing task, it would be difficult to explain the reason behind the good results. Good performance could be accounted for if the participants of the diagram group indeed constructed a good mental image of the diagram. This implies that they processed the diagram more deeply, as by noticing its part-whole relationship, the continuous relations of components, and so on, by comparing the original diagram and their drawing diagrams. Alternatively, had if they are allowed to see the original diagram during the drawing task, they might process the diagram more superficially because they need not invest much cognitive effort in the drawing task but only imitate the original diagram to reproduce it. Therefore, it is better to not allow the diagram group to see the original diagram during the drawing task.

Data Selection and Scoring Criteria

Eye-movement data from three participants were discarded due to apparent drift or missing data. Thus, data from 39 participants (20 from the text group and 19 from the diagram group) were included in the analysis.

The reading tests were also analyzed. All correct answers in the reading tests were verified by a mechanics professor. The scoring criteria for the comprehension tests were as follows: (1) For the connective relations of components (measuring partial connecting relations), participants were awarded 1 point to draw each connective relation of two components correctly and write their labels. For example, if a participant drew the handle connected to the connecting rod and wrote both labels on the picture, the total score for the connective relations of the components was 6 points. (2) For the location of components relative to the flushing cistern (measuring the part-whole relations), participants were awarded 1 point to draw the position of each component relative to the flushing cistern correctly. For example, the siphon bell must be drawn in the appropriate location with some distance from the bottom side of the flushing cistern; otherwise, the flushing cistern could not operate in principle. The total score of the location of the components relative to the flushing cistern was 10 points. (3) For similarity, two raters independently rated the similarity between the intended picture (Hegarty et al., 2003) and the picture participants drew on a five-point scale (1 = *very different* to 5 = *very similar*). Inter-rater reliability was .87, and the two raters discussed and resolved any inconsistencies until they reached a consensus.

To address the research question of how the spatial representation that readers constructed influenced the kinematic text reading, we only analyzed the eye movements during the kinematic text reading, but not during the spatial configuration text or diagram reading. We first defined the areas of interest (AOIs) as the whole text; the outlet-process sentences (sentences 2–4), the inlet-process sentences (sentences 5–7), and the target words of the 10 components in the flushing cistern were written in the kinematic text (see Figure 2). The global inspection (the whole text and sentences on the outlet and inlet systems) of eye-movement indicators related to the existing article-reading research (e.g., Ariasi et al., 2017; Jian, 2021), along with the total fixation duration (the sum of all fixations on the whole text or the sentences, which provides an indication of the overall difficulty and cognitive demand of a given reading material), the first-pass reading time (the sum of all fixations on a sentence prior to leaving the sentence, which represents the initial processing in comprehending the sentence), and the second pass (or rereading) reading time (the sum of all fixations excluding the first-pass reading time, which reflects the reanalysis of the sentence due to a lack of comprehension or in order to integrate text information). The local inspection (the target words of the component names in the flushing cistern) of eye-movement indicators used in this study related to the existing research on word identification (Kaakinen et al., 2003), along with total fixation duration (the sum of all fixations on the whole text or the sentences, as defined above), gaze duration (the sum of all fixations on a word prior to leaving it, which reflects the reading process of word recognition), and rereading time (the sum of all fixations in which readers go back to read a word after leaving it, which reflects the uncertainty or incomprehension of the word meanings).

Results

Learning Outcomes

To compare the learning outcomes of the text and diagram groups and the performance on the two tests, a two-way mixed design ANOVA was conducted with the two groups (text and diagram) as a between-subject variable, and the two tests (first and second) as a within-subject variable, with test scores as independent variables. The means and standard deviations for the spatial configuration test are shown in Table 1.

The Yes-or-no Questions

The analysis of the *yes-or-no spatial configuration questions* found significant main effects of group, $F(1, 37) = 13.64$, $p < .01$, $\eta^2 = .27$, and of the two tests, $F(1, 37) = 44.37$, $p < .001$, $\eta^2 = .55$. The interaction between groups and tests was also significant, $F(1, 37) = 11.13$, $p < .01$, $\eta^2 = .23$. Simple effects tests showed that the accuracy was significantly greater in the text group than in the diagram group on the first test, $F(1, 37) = 19.54$, $p < .001$, $\eta^2 = .33$, but not in the second test, $p > .05$. Both groups were more accurate on the second test than on the first test, $F(1, 19) = 8.39$, $p < .01$, $\eta^2 = .31$, and $F(1, 18) = 36.27$, $p < .001$, $\eta^2 = .67$, respectively.

The Drawing Test

The analysis of the drawing spatial configuration test, as for the scores of the continuous relations of components, found significant main effects of groups, $F(1, 37) = 32.85$, $p < .001$, $\eta^2 = .47$, and tests, $F(1, 37) = 12.48$, $p < .01$, $\eta^2 = .25$. The interaction between groups and tests was also significant, $F(1, 37) = 12.48$, $p < .01$, $\eta^2 = .25$. Simple effects tests showed that the accuracy was significantly greater in the text group than in the diagram

Table 1
Accuracy on the Spatial Tests for Text Group and Diagram Group

Spatial tests	Text group	Diagram group
Yes-no questions (%)		
First test	88(17)	68(10)
Second test	95(10)	89(12)
Drawing question		
Partial relations		
Continuous relations of components (%)		
First test	98(5)	75(17)
Second test	98(5)	88(11)
Locations of part-and-whole relations (%)		
First test	91(10)	85(14)
Second test	92(10)	87(13)
Global similarity (5 points)		
First test	2.15(.88)	3.26(.73)
Second test	2.20(.83)	3.37(.76)

group on both tests, $F(1, 37) = 33.13$, $p < .001$, $\eta^2 = .47$, and $F(1, 37) = 15.42$, $p < .001$, $\eta^2 = .29$, respectively. The text group showed no significant differences in accuracy between the two tests ($p > .05$); however, the diagram group was more accurate on the second test than the first, $F(1, 18) = 13.57$, $p < .01$, $\eta^2 = .43$. The scores of the component locations in part-whole relations showed no main effects of groups or tests, *p*s > .05, nor was there an interaction between groups and tests, $p > .05$. The scores for global similarity showed a main effect of group, $F(1, 37) = 20.19$, $p < .001$, $\eta^2 = .35$, such that the diagram group drew pictures more similar to the intended picture than did the text group in both tests, $F(1, 37) = 18.43$, $p < .001$, $\eta^2 = .33$, and $F(1, 37) = 20.84$, $p < .001$, $\eta^2 = .36$, respectively. There was a marginally significant main effect of tests, $F(1, 37) = 3.17$, $p = .083$, $\eta^2 = .08$, but there was no interaction between groups and tests ($p > .05$). An example of participants' drawings is shown in Figure 3.

Eye-movement Analysis

To understand how the first-stage learning of spatial configuration influences the second-stage learning of kinematic information, specifically with reference to detailed text content (outlet process and inlet process sentences), we analyzed readers' eye movements while reading the kinematic text using *t*-tests. The dependent measures were several eye-movement indicators: total fixation duration, first-pass reading time, second-pass reading time, gaze duration, and rereading time. The global inspection (the whole text and sentences on outlet and inlet systems) and local inspections (target words of the component names in the flushing cistern) of the eye movements are reported in Table 2.

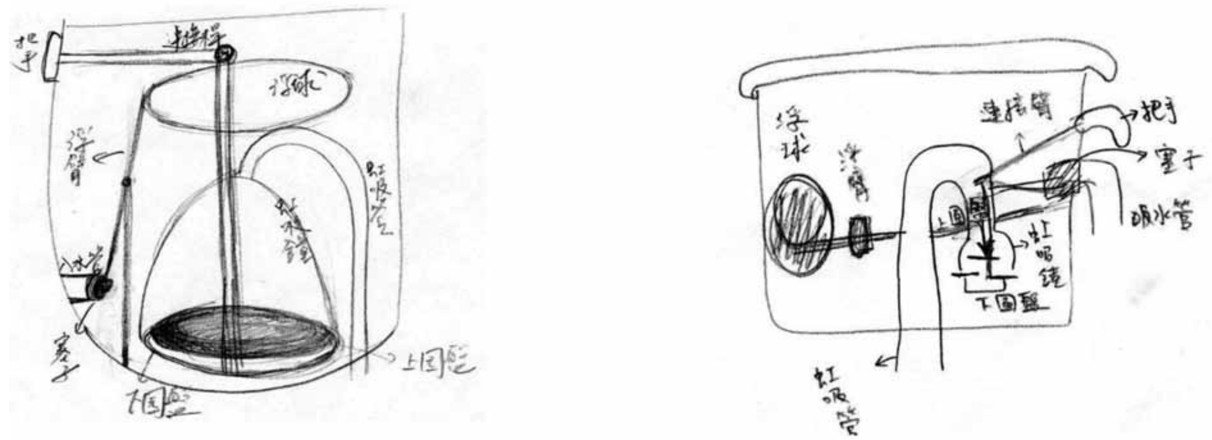
The Expository Texts

We found that the eye-movement patterns of reading expository texts showed no significant differences between the spatial and diagram groups, including total fixation duration, average fixation duration, and mean saccade length, *p*s > .05. Therefore, we conducted tests to ensure that the differences in the following eye movement analyses of the kinematic text reading were not due to the reading ability of expository texts.

Figure 3

Example drawings. (A) is a drawing by a participant in the text group, and (B) is a drawing by a participant in the diagram group

(A) (B)



The Text of Kinematic Information

For the global inspection, the results of this study showed that the text group had significantly shorter total reading times on the whole text and outlet-process sentences than the diagram group, $t(37) = -2.35$, $p < .05$, $d = .75$ and $t(37) = -2.23$, $p < .05$, $d = .71$, respectively. The text group had marginally significantly shorter second-pass reading times on the outlet-process sentences than the diagram group, $t(37) = -1.87$, $p = .069$, $d = .60$. However, there was no significant between-group difference in first-pass reading time on the outlet-process or inlet-process sentences (ps > .05).

Regarding local inspection, the text group had shorter total fixation durations ($t(37) = -2.77$, $p < .01$, $d = .88$) and shorter rereading times ($t(37) = -2.89$, $p < .01$, $d = .92$) on the target words of outlet-process operation than the diagram group. However, the two groups showed no significant differences in total fixation duration and rereading time on the target words of the inlet-process operation, ps > .05. Finally, for the eye-movement indicators of gaze durations, which reflect word identification, the groups did not significantly differ, ps > .05, whether for the target words of outlet- or inlet-process operations.

Discussion

Spatial configuration is related to kinematic operation when constructing a mental model for comprehending machine operations (Hegarty, 1992; Hegarty & Just, 1993). Many previous studies have examined how learners construct kinematic representations from words and pictures (Boucheix & Lowe, 2010; Hegary, 1992; Hegarty & Just, 1993; Heiser & Tversky, 2006; Jian & Wu, 2016; Jian et al., 2014), this study aimed to investigate how words and pictures convey spatial configurations in a machine reading condition, as well as the relations between spatial and kinematic representations.

First, we confirmed the first research question, that readers were capable of constructing the spatial configuration of a machine system by reading words, and that words and pictures had different functions in conveying different dimensions of the spatial configuration. This conclusion was supported by two sets of data (yes-no question, drawing test): the text group had better performance on the yes-or-no spatial configuration questions and on the connective relations of components in the drawing test than the diagram group. The text group had 98% accuracy on the scoring dimension of connective relations of components, and 91% accuracy on the scoring dimension of the locations of components relative to the

Table 2
Means and Standard Deviations for Eye-Movement Measures for Text group and Diagram Group on the Kinematic Text

	Text group (***N* = 20**)	**Diagram group** (***N* = 19**)
The whole text (seconds)		
Total fixation durations	183.14 (76.51)	252.79 (107.17)
Sentences (seconds)		
Total fixation durations		
Outlet system	85.90 (45.29)	119.63 (49.11)
Inlet system	34.10 (16.93)	42.05 (19.66)
First-pass reading time		
Outlet system	30.50 (13.99)	36.53 (21.68)
Inlet system	14.25 (6.77)	13.95 (6.65)
Second-pass reading time		
Outlet system	55.40 (44.45)	83.05 (47.92)
Inlet system	20.05 (17.34)	27.89 (18.67)
Words (milliseconds)		
Total fixation durations		
All system	1743.15 (841.08)	2566.40 (1026.43)
Outlet system	2130.40 (1139.28)	3346.17 (1580.12)
Inlet system	1091.25 (635.09)	1399.76 (622.84)
Number of fixations		
All system	6.86 (2.94)	9.93 (4.12)
Outlet system	8.28 (4.03)	12.92 (6.45)
Inlet system	4.43 (2.31)	5.43 (1.94)
Gaze durations		
All system	339.58 (108.40)	333.60 (78.77)
Outlet system	381.59 (161.94)	330.55 (124.43)
Inlet system	273.46 (98.00)	331.64 (122.21)
Rereading time		
All system	1403.57 (820.86)	2232.80 (1045.25)
Outlet system	1748.80 (1123.00)	3015.62 (1586.59)
Inlet system	817.79 (628.01)	1068.11 (629.19)

whole machine. The findings were consistent with the conclusion of Schnotz et al. (2014), who suggested that text and pictures have different functions in knowledge acquisition and cognitive processing. As in previous research (Schmidt-Weigand & Scheiter, 2011), our study showed that undergraduate readers were able to construct spatial representations from reading pure text (the steps of lightning formation) describing spatial information.

It also indicated that mature readers were capable of constructing a spatial representation by reading words describing the spatial configuration of a flushing cistern and accurately captured the connective relations of its components. This indicated strong word superiority in conveying detailed relations and confirmed previous suggestions that verbal descriptions are also effective in conveying spatial structure (Mani & Johnson-Laird, 1982; Medin et al., 2005; Taylor & Tversky, 1992). There are two possible explanations for the participants' strong performance in constructing the spatial mental image by reading textual descriptions. First, they were mature adult readers with rich experience in imagining pictures by reading words or hearing verbal information. The second reason pertains to the linguistic characteristics of Chinese people. There were many important terms in the reading material used in this study, carrying visual-spatial information. For example, 上圓盤 is a Chinese phrase consisting of three characters, *upper-circle-plate*; 浮球 'float ball' provides spatial and shape information. Adult readers can draw some characteristics of the components based on the meanings of the component characters of the names.

In contrast, the diagram group performed better on the spatial test measuring overall similarity. The appearance of the drawings of the diagram group better approximated the intended picture than did the drawings of the text group, but the detailed relations of the components they constructed were not precise enough. This result might suggest that the information in the diagram is grouped in visualizations (Goldstone & Son, 2005; Schmidt-Weigand & Scheiter, 2011), so it was easier to access its overall appearance from memory. This can also be explained based on the processing route of picture comprehension in the *model of text and picture comprehension* (Schnotz & Bannert, 2003), whereby an individual creates a visual representation of the pictures through perceptual processing, which organizes graphic entities in the visuo-spatial sketchpad in working memory according to Gestalt Laws (Ohlsson, 1984; Schnotz & Wagner, 2018). However, detailed information on the overall mental image may not be sufficiently precise. In contrast, examination of the pictures that participants drew showed that the text group readers drew many correctly detailed connective relations of components, but the picture was unlike the intended picture as a whole; in particular, the components' locations, sizes, and relative distances were very different from the intended picture (see Figure 3).

Moreover, we confirmed the second research question about the interaction relations between spatial configuration and kinematic representation. This result provides direct empirical evidence supporting the perspective of previous research (Heiser & Tversky, 2006), indicating that constructing the kinematic representation of a machine operation must be based on its spatial configuration. As predicted by Hypothesis 2a, the results indicated that different internal spatial representations of the spatial and diagram groups influenced the subsequent kinematic text reading. The two groups had different eye-movement patterns, especially for sentences describing complicated concepts. We found that the text group had significantly shorter total reading time and rereading time on the outlet-process sentences and component words than the diagram group. However, the groups showed no differences in eye movements when processing the simple concept of inlet-process sentences and their component words. Because second-pass reading time reflects reanalysis of the sentences due to lack of comprehension or in order to integrate the textual information (Ariasi et al., 2016; Jian et al., 2019; Mason et al., 2013), the spatial representations constructed by the diagram group were not as easily integrated into the kinematic representation in the following reading, as in the text group. The diagram group readers captured limited continuous relations of components (evidenced by the drawing test score) in the first-stage reading on viewing the flushing cistern diagram, so they compensated for this spatial representation construction while reading the kinematic text in the second-stage reading, resulting in a longer processing time on the kinematic text. In summary, the above data indicate that to comprehend complicated kinematic concepts, it is more helpful to use words to represent spatial configuration, while for comprehending simple kinematic concepts, words and diagrams have similar benefits.

Moreover, the results of this study confirmed some of the predictions of Hypothesis 2b. Both groups had higher accuracy on the yes-no questions in the second spatial test than the first, while in the drawing test, the diagram group was more accurate on the second test than on the first test in drawing the continuous relations of components. However, neither group improved significantly in drawing the correct component locations of part-and-whole relations in the second drawing test, while for global similarity on the drawing test, both groups showed a tendency to improve that only reached marginal significance. In summary, the above data indicate that the kinematic information of a machine operation described in words would allow for the revision of some parts of readers' inner spatial configuration representations, especially of the dimension of continuous relations of components.

Contributions and Limitations

This study makes several contributions to the literature. First, this study provides empirical data reflecting learning outcomes (spatial tests) and learning processes (eye movements) to verify the model of text and picture comprehension (Schnotz & Bannert, 2003), and extends it to demonstrate the nature of mental models created by readers when visual inputs are descriptive or depictive representations. Second, this study refined our knowledge of the contents of spatial configuration and provided empirical evidence that words and pictures have different functions in conveying them. Pictorial representations have the advantage of conveying global characteristics, but verbal representations can better convey the characteristics of the detailed relations of the connections in the spatial configuration of a machine system. It is worth noting that whether the measurement was of pictorial (drawing a picture) or verbal representations (yes-or-no questions), the participants who read verbal descriptions of spatial configurations captured more detailed relations of the components than those who read the machine diagram. Third, this study indicates which content of the spatial configurations readers constructed can (continuous relations of components) and cannot (part-whole relations) be revised by the kinematic representation. Fourth, the findings indicated that a well-designed text should provide the necessary information for readers to construct the mental spatial representation of a machine system. The necessary information should at the least includes components' names, nouns of locality, verbs of alteration of direction, and so forth. In addition, text difficulty and readers' abilities must also be taken into consideration.

Some limitations of this study need to be considered. First, this study recruited undergraduate students as participants. They were mature readers, and the 10 component labels were Chinese words with high semantic transparency (the word's meaning is easily decoded from its constituent characters' meanings). Therefore, these participants were capable of decoding the visual–spatial information of Chinese words through their constituent characters and were able to draw some components' spatial characteristics accordingly. However, this result might not be generalizable to young readers whose reading abilities do not allow them to decode words' meanings without overt teaching. Furthermore, if the component labels were semantically opaque words (word meanings cannot be decoded based on the meanings of the constituent characters), readers might not have been able to construct their spatial representations by reading the labels. Teachers should recognize these situations and evaluate the words' properties and text difficulty for their young students while teaching them to construct a spatial representation during text reading. Future research can account for developmental issues and recruit young readers to compare their results with those of the adult readers in this study. Scholars could also examine whether the results can be generalized to alphabetic writing systems. Second, the drawing task in this study allowed participants to draw and read the spatial description text simultaneously, which might lead to overestimations of the text group's learning outcomes from reading the spatial text, because they read the spatial description text twice. By contrast, if we asked participants to draw the machine's configuration without providing the original text describing the spatial configuration of the flushing cistern, the results might underestimate their performance. As previously mentioned, we adopted the former method out of an abundance of caution: if we had not provided the original text for readers and they had not drawn the spatial configuration very well, we would not have been able to distinguish whether the poor performance was due to poor memory or an inability to construct a spatial mental image from reading the verbal description. A future study may consider adding another experimental group who are asked to complete the drawing task without looking at the spatial description text, and to examine if the three groups had different test performance. Third, the reading material used in this study only focused on the machine configuration. Therefore, the three dimensions of the proposed spatial representation may not cover other domains. For example, in geography and astronomy, distance is a very important dimension of spatial representation, but we did not consider it in our classification. Moreover, this study only used yes-no questions and a drawing test to measure readers' spatial representations. An operation test may also be included in further studies to investigate the multidimensionality of spatial representation. These research limitations and suggestions shed light on the future directions of spatial representation research.

References

Ainsworth, S., & VanLabeke, N. (2004). Multiple forms of dynamic representation. *Learning and Instruction*, *14*(3), 241-255. https://doi.org/10.1016/j.learninstruc.2004.06.002

Ariasi, N., Hyönä, J., Kaakinen, J. K., & Mason, L. (2017). An eye-movement analysis of the refutation effect in reading science text. *Journal of Computer Assisted Learning*, *33*(3), 202-221. https://doi.org/10.1111/jcal.12151

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511527685

Boucheix, J.-M., & Lowe, R. K. (2010). An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learning and Instruction*, *20*(2), 123-135. https://doi.org/10.1016/j.learninstruc.2009.02.015

Eitel, A., Scheiter, K., Schüler, A., Nyström, M., & Holmqvist, K. (2013). How a picture facilitates the process of learning from text: Evidence for scaffolding. *Learning and Instruction*, *28*, 48-63. https://doi.org/10.1016/j.learninstruc.2013.05.002

Goldstone, R. L. & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the Learning Sciences*, *14*(1), 69-110. https://doi.org/10.1207/s15327809jls1401_4

Greeno, J. G., & Simon, H. A. (1974). Processes for sequence production. *Psychological Review*, *81*(3), 187-198. https://doi.org/10.1037/h0036340

Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *18*(5), 1084-1102. https://doi.org/10.1037//0278-7393.18.5.1084

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, *32*(6), 717-742. https://doi.org/10.1006/jmla.1993.1036

Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognition and Instruction*, *21*(4), 209-249. https://doi.org/10.1207/s1532690xci2104_1

Heiser, J., & Tversky, B. (2006). Arrows in comprehending and producing mechanical diagrams. *Cognitive Science*, *30*(3), 581-592. https://doi.org/10.1207/s15516709cog0000_70

Jian, Y.-C. (2016). Fourth graders' cognitive processes and learning strategies for reading illustrated biology texts: Eye movement measurements. *Reading Research Quarterly*, *51*(1), 93-109.

Jian, Y.-C. (2021). The immediate and delayed effects of text-diagram reading instruction on reading comprehension and learning processes: evidence from eye movements. *Reading and Writing*, *34*, 727-752. http://dx.doi.org/10.1007/s11145-020-10089-3

Jian, Y.-C., & Ko, H.-W. (2017). Influences of text difficulty and reading ability on learning illustrated science texts for children: An eye movement study. *Computers and Education*, *113*, 263-279. https://doi.org/10.1016/j.compedu.2017.06.002

Jian, Y.-C., Su, J.-H., & Hsiao, Y.-R. (2019). Differentiated processing strategies for science reading among sixth-grade students: exploration of eye movements using cluster analysis. *Computers and Education*, *142*. 103652. https://doi.org/10.1016/j.compedu.2019.103652

Jian, Y.-C., & Wu, C.-J. (2016). The function of diagram with numbered arrows and text in helping readers construct kinematic representations: Evidenced from eye movements and reading tests. *Computers in Human Behavior*, *61*, 622-632. https://doi.org/10.1016/j.chb.2016.03.063

Jian, Y.-C., Wu, C.-J., & Su, J.-H. (2014). Learners' eye movements during construction of mechanical kinematic representations from static diagrams. *Learning and Instruction*, *32*, 51-62. https://doi.org/10.1016/j.learninstruc.2014.01.005

Kaakinen, J. K., Hyona, J., & Keenan, J. M. (2003). How prior knowledge, WMC, and relevance of information affect eye fixations in expository text. *Journal of Experimental Psychology: Learning*, *Memory*, *and Cognition*, *29*(3), 447-457. http://dx.doi.org/10.1037/0278-7393.29.3.447

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363-394. https://doi.org/10.1037/0033-295X.85.5.363

Kriz, S., & Hegarty, M. (2007). Top-down and bottom-up influences on learning from animations. *International Journal of Human-Computer Studies*, *65*(11), 911-930. https://doi.org/10.1016/j.ijhcs.2007.06.005. https://doi.org/10.1016/j.ijhcs.2007.06.005

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*(1), 65-100. http://doi.org/10.1016/S0364-0213(87)80026-5

Libben, G., Gibson, M., Yoon, Y. B., & Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language*, *84*(1), 50-64. http://doi.org/10.1016/S0093-934X(02)00520-5

Mani, K., & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory and Cognition*, *10*, 181-187. https://doi.org/10.3758/BF03209220

Mason, L., Pluchino, P., Tornatora, M. C., & Ariasi, N. (2013). An eye-tracking study of learning from science text with concrete and abstract illustrations. *Journal of Experimental Education*, *81*(3), 356-384. https://doi.org/10.1080/00220973.2012.727885

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, *82*(4), 715-726. https://doi.org/10.1037/0022-0663.82.4.715

McCrudden, M. T., Magliano, J. P., & Schraw, G. (2011). The effect of diagrams on online reading processes and memory. *Discourse Processes*, *48*(2), 69-92. https://doi.org/10.1080/01638531003694561

Medin, D., Ross, B. H., & Markman, A. B. (2005). *Cognitive psychology* (4th ed.). John Wiley & Sons.

Ohlsson, S. (1984). Restructuring revisited: I. Summary and critique of the Gestalt theory of problem solving. *Scandinavian Journal of Psychology*, *25*(1), 65-78. https://doi.org/10.1111/j.1467-9450.1984.tb01001.x

Schmidt-Weigand, F., & Scheiter, K. (2011). The role of spatial descriptions in learning from multimedia. *Computers in Human Behavior*, *27*(1), 22-28. https://doi.org/10.1016/j.chb.2010.05.007

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representations. *Learning and Instruction*, *13*(2), 141-156. https://doi.org/10.1016/S0959-4752(02)00017-8

Schnotz, W., Ludewig, U., Ullrich, M., Horz, H., McElvany, N., & Baumert, J. (2014). Strategy shifts during learning from texts and pictures. *Journal of Educational Psychology*, *106*(4), 974-989. https://doi.org/10.1037/a0037054

Schnotz, W., & Wagner, I. (2018). Construction and elaboration of mental models through strategic conjoint processing of text and pictures. *Journal of Educational Psychology*, *110*(6), 850-863. https://doi.org/10.1037/edu0000246

Taylor, H. A., & Tversky, B. (1992). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language*, *31*, 261-282. https://doi.org/10.1016/0749-596X(92)90014-O

Tsai, M.-J., Wu, A.-H., & Chen, Y. (2019). Static and dynamic seductive illustration effects on text-and-graphic learning processes, perceptions, and outcomes: Evidence from eye tracking. *Applied Cognitive Psychology*, *33*(1), 109-123. https://doi.org/10.1002/acp.3514

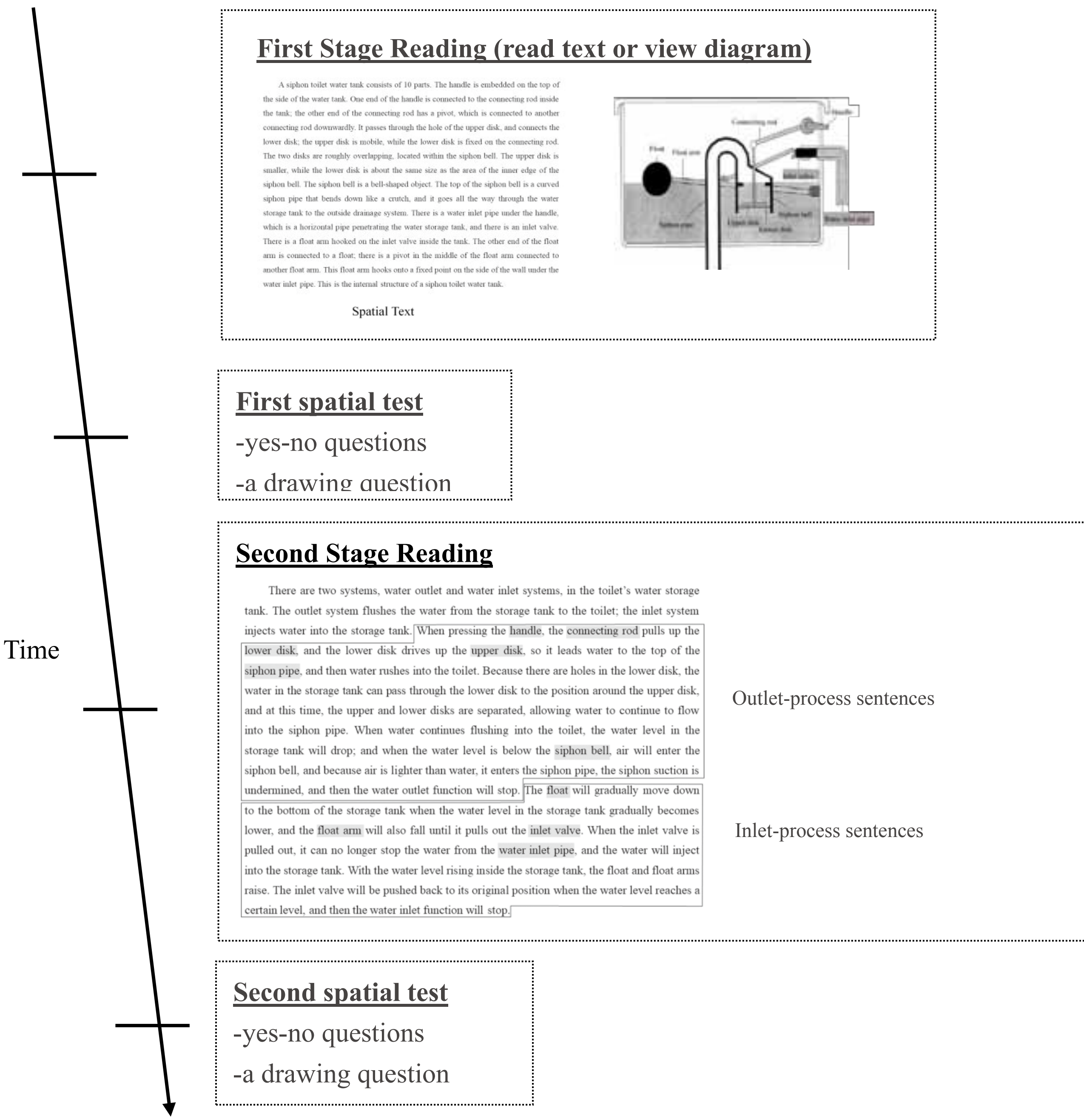
Tversky, B. (2001). Spatial schemas in depictions. In M. Gattis (Ed.), *Spatial schemas and abstract thought* (pp.79-111). The MIT Press.

Wu, C.-J., & Liu, C.-Y. (2021). Eye-movement study of high- and low-prior-knowledge students' scientific argumentations with multiple representations. *Physical Review Physics Education Research*, *17*(1), 010125. https://doi.org/10.1103/PhysRevPhysEducRes.17.010125

Wu, C.-J., Liu, C.-Y., Yang C.-H., & Wu C.-Y. (2021). Children's reading performances in illustrated science texts: Comprehension, eye movements, and interpretation of arrow symbols. International Journal of Science Education, *43*(1), 105-127. https://doi.org/10.1080/09500693.2020.1858515

Appendix

Figure 2
(English version). The experimental procedures and materials



DOI:10.6129/CJP.202203_64(1).0003
中華心理學刊 民 111，64 卷，1 期，57-76
Chinese Journal of Psychology 2022, Vol.64, No.1, 57-76

閱讀文字和圖片以建構空間結構和動態表徵的眼動研究

簡郁芩、吳昭容
國立臺灣師範大學教育心理與輔導學系

本研究探討讀者如何透過閱讀文字和圖片以建構空間結構和動態表徵，以及此兩種表徵之間的關係為何。本實驗為兩階段設計：在第一階段，大學生受試者被隨機分派到文字組（閱讀描寫儲水槽空間結構的文本）或圖片組（看一張儲水槽結構的圖片），然後，完成一個空間結構測驗；在第二階段，兩組讀者閱讀一篇相同的文章，內容描述儲水槽的動態運作歷程，並讓受試者修改他們第一次空間測驗的答案。受試者閱讀動態訊息文章時的眼動型態會被記錄下來。研究結果顯示，文字組在空間結構測驗中畫出儲水槽內各個部件（如：把手、連接桿、上下圓盤等）之間的關係較正確，代表文字較有助於讀者形成部件之間的接續關係；但圖片組畫出來的圖與原圖相似度較高，代表圖片較有助於讀者形成整體結構的空間表徵。而從第二次修改測驗的結果發現，讀者在讀完動態訊息文本之後所形成的動態表徵主要是幫助讀者修改空間表徵中的「部件接續關係」，而非「部件與整體的位置關係」。此外，眼動資料也顯示讀者從文字與圖片所形成的空間表徵有所不同，且空間表徵會影響讀者在後續動態訊息文本的閱讀歷程：文字組讀動態訊息文本的總凝視時間和重讀時間都較圖片組短，特別是在動態訊息文本中概念較不複雜的句子（描述儲水槽的出水運作歷程）和儲水槽部件的目標詞上。綜上，本研究確認文字和圖片於傳達空間結構訊息的功能有所不同，亦說明讀者在閱讀過程中形成空間表徵與動態表徵的互動關係，進而建構文章內容的心智模式。

關鍵字：眼動型態、動態表徵、閱讀、空間結構