International Journal of Science Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tsed20

Children's reading performances in illustrated science texts: comprehension, eye movements, and interpretation of arrow symbols

Chao-Jung Wu, Chia-Yu Liu, Chung-Hsuan Yang & Chen-Ying Wu

To cite this article: Chao-Jung Wu, Chia-Yu Liu, Chung-Hsuan Yang & Chen-Ying Wu (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, DOI: 10.1080/09500693.2020.1858515

To link to this article: https://doi.org/10.1080/09500693.2020.1858515

Published online: 14 Dec 2020.

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION
2021, VOL. 43, NO. 1, 105–127
https://doi.org/10.1080/09500693.2020.1858515

Routledge
Taylor & Francis Group

Check for updates

Children's reading performances in illustrated science texts: comprehension, eye movements, and interpretation of arrow symbols

Chao-Jung Wu, Chia-Yu Liu, Chung-Hsuan Yang and Chen-Ying Wu

Department of Educational Psychology and Counseling/Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taipei, Taiwan

ABSTRACT

The study aims to explore the effect of illustrations on young readers' comprehension of science passages, including how they decode specific symbols such as arrows when viewing illustrations. We investigated reading behaviours and interpretation of arrow symbols of 64 sixth-graders in three illustrated science passages using an eye tracker and by conducting tests and interviews. Results showed no significant difference between the illustrated and text-only groups on reading comprehension, total fixation duration for the text section, and FD/P ratio (the total fixation duration for each paragraph divided by their area [numbers of pixels]) of the paragraphs with spatial structure information. Furthermore, the average duration of the illustrated group's fixation upon illustrations was less than ten seconds. Regarding reading sequence, the illustrated group's referencing behaviour between text and illustrations were limited and had inappropriate timing. Participants could be aware of the semantic roles of the arrows in major categories but had difficulty distinguishing subcategories. Furthermore, their use of the semantic role of 'labelling' was overextended and that of 'vector' was underextended. The words that young readers use most frequently to refer to the subcategories signified by the arrow symbols were identified. The implications of instructions for understanding diagram conventions are discussed.

ARTICLE HISTORY
Received 20 May 2020
Accepted 29 November 2020

KEYWORDS
Arrow symbols; eye-tracking methodology; illustrations; young readers

Introduction

Multiple representations composed of illustrations and written text have been widely used in scientific journals, textbooks, and online information (Lee, 2010; Liu & Khine, 2016; Smith & Pol, 2018). Bowen and Roth (2002) confirm that illustrations played a dominant role in scientific texts and lectures. Past research has confirmed that adults learn more effectively when illustrations and texts are provided simultaneously rather than texts alone (Schweppe et al., 2015), for pictures can be used as a mental scaffolding

CONTACT Chia-Yu Liu leave1756@gmail.com Department of Educational Psychology and Counseling/Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, No. 162, Sec. 1, Heping E. Rd., Taipei 106, Taiwan

to improve reading comprehension and facilitate the processing of the text (Eitel et al., 2013). However, the debate about the effects of illustrations on young readers remains. Some studies have shown that illustrations improved students' learning (Jian & Ko, 2017; Mason, Pluchino, et al., 2013), but others failed to find such an effect (Elen & Gorp, 2008; Hannus & Hyönä, 1999). In various studies pertaining to the comprehension of text and illustration, eye movement data were adopted to understand individuals' priority of judging certain areas and how they inspect text and illustrations by transitioning their fixations between these areas (Alemdag & Cagiltay, 2018; Jian, 2016; Mason, Tornatora, et al., 2013; Rayner et al., 2001). Adult readers demonstrated consistent referencing behaviours between the text and illustrations (Rayner et al., 2001). However, other studies found that young readers switched between pictorial and textual information (Mason, Pluchino, et al., 2013), or made limited connections between the two representations (Jian, 2016). Thus, the first aim of this study is to clarify the effect of using illustrations in scientific texts on the reading comprehension and eye movements of young readers.

Decoding arrow symbols is critical for readers to comprehend illustrated science texts. Previous studies have indicated that arrow symbols might be beneficial for undergraduate and high school students' learning outcomes (Cromley, Bergey, et al., 2013; Cromley, Perez, et al., 2013; Jian & Wu, 2016). In addition, arrows are semantical symbols commonly used in scientific diagrams (Coleman et al., 2011). They are versatile symbols that can convey complicated causal relationships with precise information (Kurata & Egenhofer, 2008). In the past, very few studies have investigated how young readers comprehend the meaning of arrow symbols in illustrated science passages. Thus, the second aim of this study is to examine whether young readers comprehend the semantic roles of arrow symbols and understand how they explain the arrow symbols in their language.

Theory of text and picture comprehension

In recent years, studies have revealed that reading comprehension can be improved by the integration of texts and illustrations rather than texts alone (Levie & Lentz, 1982). The potential benefits of multimedia materials that comprise texts and illustrations are usually explained by the Cognitive Theory of Multimedia Learning (CTML; Mayer, 2009) or the Integrative Model of Text and Picture Comprehension (ITPC; Schnotz, 2014). According to CTML, individuals select relevant texts and pictures from the materials, organise a text-based model and a picture-based model separately, and integrate them into a coherent mental representation by using the readers' prior knowledge. According to ITPC, individuals first generate a representation based on the surface structure of the text, and then develop a propositional representation according to the semantic content. Subsequently, a visual representation based on the visual image is generated and mapped onto semantic relations to offer structure for the mental model. Finally, a coherent mental model is formed by the constant interactions between the mental model and the propositional representation. Both CTML and ITPC confirm the interaction effect between picture and text at various levels.

Eitel et al. (2013) proposed the scaffolding assumption and further examined the different effects of text and pictures that how they assist individuals' construction of mental model. They proposed that text is appropriate to convey information on a general proposition level and picture is well suited to provide structure-related information of a system (e.g. the spatial structure of a pulley system). Thus, the more ambiguous a text is about the structure information, the more benefits a corresponding picture could provide. They found that comprehension fostered if the participants inspected the picture as reading an ambiguous text rather than reading only the text. It indicated that pictures act as a mental scaffold, which constrain the interpretations from text that contains unclear structure information, in turn fostering comprehension. Another finding showed that participants who inspected the picture before the text spent shorter time on the structure-related section of the text than participants who read only the text. It revealed that a picture is beneficial for constructing a mental model about the object's spatial structure, and then facilitating processing of structure-related text, thus demanding shorter reading time. However, the participants Eitel et al.'s study were adults, and not children.

Effects of illustrations on young readers

The effectiveness of illustrations in teaching materials for young readers may not always lead to greater learning success. Several studies have shown the effectiveness of illustrations, but the effect might depend on the abilities of the young readers (Jian, 2020; Jian & Ko, 2017; Reid & Beveridge, 1986), the contents of passages (McTigue, 2009), or the features of illustrations (Mason, Pluchino, et al., 2013). For example, Jian and Ko (2017) as well as Reid and Beveridge (1986) found that illustrations were beneficial for high-ability children but inhibited the performance of low-ability children. McTigue (2009) investigated the effects of illustrations in explanatory sciences passages on the reading comprehension of middle-school students. The results indicated that the students benefitted from illustrations only in life science texts, but not in physical science texts. A study by Mason, Pluchino, et al. (2013) found that readers of texts with labelled illustrations in science text showed better performance on a transfer test than those with unlabelled illustrations and only text.

On the other hand, some studies have found that inclusion of illustrations offer no advantages (Hannus & Hyönä, 1999) or have limited effects (Elen & Gorp, 2008) on young readers. In a study by Hannus and Hyönä (1999), children did not benefit from illustrations in their ability to recall the main points of a text. Elen and Gorp (2008) investigated the effect of multimedia design features on the retention and transfer ability of 240 primary school children under 24 conditions, including the relationships between texts and pictures, image modalities, learner-control types, and materials. The results showed that the integrated effects of texts and pictures were inconsistent, and their interrelationships were complicated.

The reading sequence of young readers remained a controversial issue (Jian, 2016; Jian & Ko, 2017; Mason, Tornatora, et al., 2013). The reading sequence is defined as the readers' referencing behaviour between texts and pictures as processing illustrated science passages. Mason, Tornatora, et al. (2013) classified young readers' eye-movement patterns by cluster analysis and identified three integrating levels of text and picture. They found that the greater integration of the illustrated text lead to higher performance of the students. Comparing with adult readers, Jian (2016) indicated that young readers made fewer references between texts and illustrations. Most young readers tended to refer from the illustrations to the texts in a unidirectional way, indicating that they tend to focus on a single representation rather than multiple. However, these two studies did not examine further whether young readers could inspect pictures as closely as reading the structure section of the text.

Several eye-movement indicators were adopted to examine readers' cognitive process, including the total reading time on certain stimuli, the distribution of fixation duration, and the reading sequence (Hannus & Hyönä, 1999; Jian, 2016; Jian et al., 2014; Rayner et al., 2001). For example, Hannus and Hyönä (1999) examined readers' cognitive efforts on texts and illustrations by their fixation duration on texts and illustrations separately. In addition, Rayner et al. (2001) calculated the FD/P ratio (the total fixation duration for each paragraph divided by their area [numbers of pixels]), which is a critical index of the reading ratio in certain areas. For understanding the moment-to-moment behavioural sequences, Jian (2016) and Jian et al. (2014).) applied a series of matrix calculations to examine readers' reading sequence.

Semantics of arrow symbols

Arrows are semantic symbols frequently used in the diagrams of illustrated science passages. An arrow diagram contains at least one arrow symbol and a component, which could be an icon, a text label, or a certain position in the background diagram. Tversky (2001) defined arrow symbols as lines with traits of linearity and asymmetric that can direct readers' attention and create connections between components.

The versatility and semantic roles of arrow symbols allow them to convey abundant information. Kurata and Egenhofer (2008) have developed an algorithm for computers to identify possible semantic roles of arrow symbols. It distinguishes four major categories and several subcategories of semantic roles for arrows (Figure 1):

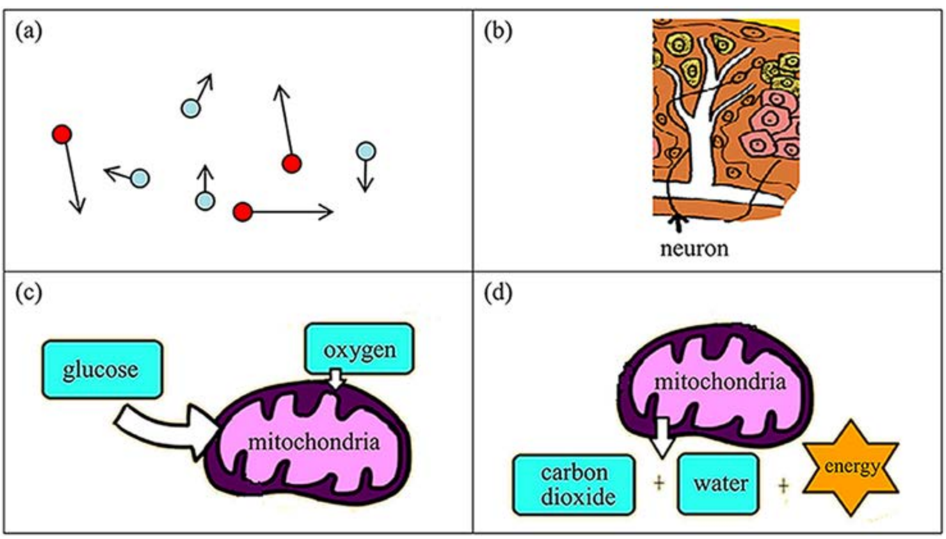

Figure 1. The semantic roles of arrow symbols: (a) Orientation: air molecules in the form of vector (b) Annotation: the label of a neuron (c) Behaviour description: the spatial movement of glucose and oxygen into mitochondria (d) Association: the change of a molecule into carbon dioxide, water, and energy through mitochondria

(1) **Orientation.** The structure of the arrow symbol for ‘orientation’ contains one subject that is used to specify the directional property. The subcategories are ‘direction’, ‘vector’, and ‘magnitude’ (e.g. moving out or passing by).
(2) **Annotation.** The structure of the arrow symbol for ‘annotation’ contains two components – one subject and one attached label. A subject is connected with a label (e.g. name or category) by an arrow symbol. The subcategory is ‘labelling’.
(3) **Behaviour description.** The structure of the arrow symbol for behaviour description includes components of one subject and one object that exist independently and can interact with each other. The subcategory is ‘spatial movement’. For example, the tail slot represents a car, and the head slot represents ‘Taipei’, indicating the movement of driving a car to Taipei.
(4) **Association.** The structure of arrow symbols for association contains components of two subjects that are linked by an arrow symbol to demonstrate a symmetrical relationship between them. The relationship between the two subjects could be logical, temporal, spatial, or imaging. The subcategories are ‘ordered relation’, ‘change’, and ‘conditional relation’. For example, the Eiffel Tower is a landmark of Paris.

Studies that have empirically investigated the effect of arrow symbols on individuals’ reading comprehension (Boucheix & Lowe, 2010; Heiser & Tversky, 2006; Jian et al., 2014; Kriz & Hegarty, 2007) have found that the most frequently used sematic roles of arrow symbols in illustrated science passages are ‘direction’, ‘labelling’, and ‘spatial movement’, which fall under the major categories of ‘orientation’, ‘annotation’, and ‘behavioural description’, respectively. Though some research has confirmed the benefits of arrow symbols for improving individuals’ reading comprehension (Heiser & Tversky, 2006; Jian et al., 2014), others have proposed that arrow symbols can only guide readers’ attention (Boucheix & Lowe, 2010; Kriz & Hegarty, 2007), indicating that the effects of arrow symbols on reading comprehension still remain controversial. This shows that the readers might not be able to recognise these semantic roles of arrow symbols correctly. Clarifying whether individuals recognise the various semantic roles of arrows and how they decipher these arrows in their words could be an approach to understand individuals’ thinking regarding various arrow symbols, which should be critical in developing future instructions for the conventions.

In order to understand individuals’ recognition performances of verbal learning and memory, signal detection theory has often been applied (Crocker et al., 2011; Kramer et al., 2005; Russo et al., 2016). This theory analysed both the individual’s hit rate (i.e. the proportion where a stimulus is present and the participant correctly responds) and the false-alarm rate (i.e. the proportion of incorrect yes responses). In this study, the hit rate, referred as the proportion of a semantic role of arrow symbol, was correctly recognised; the false-alarm rate, referred as the proportion of a semantic role of arrow symbol, was incorrectly recognised in passages. The contrast z score of hit rate and false-alarm rate yielded the discriminability index (d’). The higher d’ means better discrimination. Thus, this study adopts hit rate, false-alarm rate, and d’ to examine whether young readers recognise the semantic roles of arrows in illustrated science passages correctly. Further, this study also investigates young readers' words in interpreting the arrows' meanings.

The current study

According to Mayer (2009)'s CTML and Schnotz (2014)'s ITPC, the interplay between pictures and text happens at various processing stages. Further, Eitel et al. (2013) zoomed in by proposing the scaffolding hypothesis that inspecting a picture with structure-related information allows individuals to constrain the range of interpretations of the spatial structure from text, in turn improving the comprehension and facilitating the text processing. If young readers apply scientific illustrations as a scaffold, then their reading comprehension would be improved (**Hypothesis 1**), and the total reading time of the text sections and the time spent on the structure-related paragraphs (e.g. the space allocation for biological tissue) would be less than in young readers who read only text (**Hypothesis 2 and 3**). In addition, if young readers could utilise scientific illustrations as a scaffold, they would inspect illustrations more closely while reading the structure-related paragraphs, and perform stronger text-illustration referencing behaviours (**Hypothesis 4**).

Hypothesis 1. Reading comprehension will be better in the illustrated group.
Hypothesis 2. Total fixation duration of the text sections will be shorter in the illustrated group.
Hypothesis 3. FD/P ratio will be lower for the paragraphs with structure-related information in the illustrated group.
Hypothesis 4. The illustrated group will demonstrate more text-illustration referencing behaviours on the structure-related paragraphs.

As mentioned earlier, we suggest that investigating whether young readers can comprehend various meanings of arrow symbols and exploring the words they mostly use to interpret the arrows in scientific diagrams are necessary, for they are the basis for developing instructions to understand diagram conventions. As a result, we investigate two exploratory research questions:

Exploratory Research Question 1. Could young readers identify the various semantic roles of arrow symbols accurately?
Exploratory Research Question 2. What are the most frequently used words by the young readers to refer to the major category and subcategories of arrow symbols?

Method

Participants

Sixty-four sixth graders in northern Taiwan were screened by Ko (1999)'s Reading Comprehension Screening Test. Students who obtained parental consent and had no reading difficulty were randomly assigned to two reading groups: illustrated ($n = 34$) and text-only ($n = 30$). All participants' vision were normal or corrected-to-normal, and their ages ranged from 11.8 to 12.6 years old (55% male).

Materials

Learning and testing materials

This study contained three passages, two in the biological domain (the firefly and cellular respiration) and one in the physical domain (the hot air balloon). Passages were not taught in our participants' school curriculum. The texts and illustrations for all three passages were adapted from an encyclopaedia on science. The three illustrations in this study were modified based on the principles of Levin et al. (1987) to ensure accuracy and accessibility to average sixth graders. One elementary school teacher who holds a master's degree in science education adjusted the materials to ensure appropriate literacy level of the upper elementary students. All of the learning and testing materials were further adapted based on the results of a pilot test with five sixth graders.

Each passage was divided into four or five paragraphs (Figures 2, 3, and 4). The first paragraph introduced the topic by providing the context. The middle paragraphs explained the definition and the focus of the topic. The final paragraph provided a conceptual summary and conclusion. Since the first paragraph usually involves a brief introduction of the topic and the last paragraph contains the topic conclusion, our main focus of Hypothesis 3 and 4 are the middle paragraphs–paragraph 2 to paragraph 3 (or 4), which are more relevant with the structure-related information. All of the passages were analysed by the Chinese Readability Index Explorer (CRIE) system (Sung et al., 2015) to confirm that their textual features are similar. According to the Clay's classification

臺灣的晚春可以發現螢火蟲的蹤跡。它們生活在陰暗潮濕的環境中，喜歡在晴天、溫暖、無風的夜晚活動。

螢火蟲的光芒來自於腹部的發光器，是由發光細胞、反射細胞、氣管和神經所組成。發光細胞有螢光素和發光酵素。氣管提供了發光過程所需要的氧氣。神經則傳遞與發光有關的訊號。

螢火蟲腹部的表面有氣孔，空氣會從氣孔、經由氣管，到達發光細胞。神經傳遞訊號，通知發光細胞中的發光酵素開始作用，使螢光素和氧氣進行化學反應。這種作用所產生的能量，只有少部分變成熱，多數是以光的形式釋放。因此，螢火蟲的光被稱為「冷光」。

反射細胞可以使螢火蟲的光更明亮。如果將螢火蟲的發光器比喻成汽車的車燈，發光細胞就像是燈泡，是真正發光的部位。而反射細胞就像是反光罩，可以將發光細胞所發出的光集中反射出去，使光芒變得明亮。

螢火蟲一生只有一兩週的時間藉由發光來尋求配偶。隨著交配、產卵，完成繁殖下一代的任務，它們的生命也走到了盡頭。

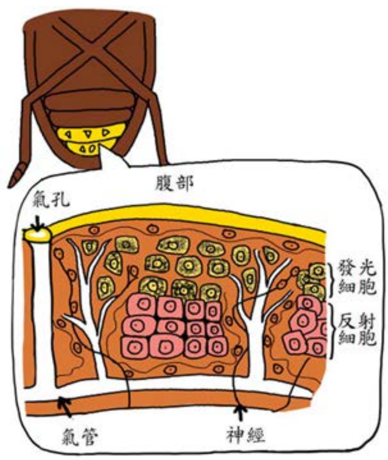

Figure 2. The illustrated passage of firefly.

呼吸作用是指細胞把體內的養分分解、產生能量，以維持生命的過程，和我們平常說的呼吸不太一樣。

呼吸作用分成兩種。一種是需要氧氣參與養分分解的過程，稱為「有氧呼吸」。另一種是「無氧呼吸」，也就是分解養分的過程中，不需要用到氧氣。

有氧呼吸是在細胞裡的粒線體進行，在分解葡萄糖時需要氧氣，分解後會產生能量、二氧化碳和水。無氧呼吸則是在細胞質裡面進行，不需要氧氣，產生的能量比較少。而無氧呼吸又可分為兩類，一類無氧呼吸最終會產生乳酸，另一類則是產生二氧化碳和酒精。

以運動而言，慢跑和快跑分別屬於有氧和無氧運動。無氧運動會產生乳酸，所以快跑之後，肌肉會感到痠痛。以水果而言，放置一陣子的蘋果會出現酒味，是因為無氧呼吸的緣故。

從產生的物質來看，有氧呼吸的產物是對生物無害的二氧化碳和水，而無氧呼吸的產物則是乳酸或酒精。過多的乳酸和酒精對生物有害，因此生物不能長時間進行無氧呼吸。

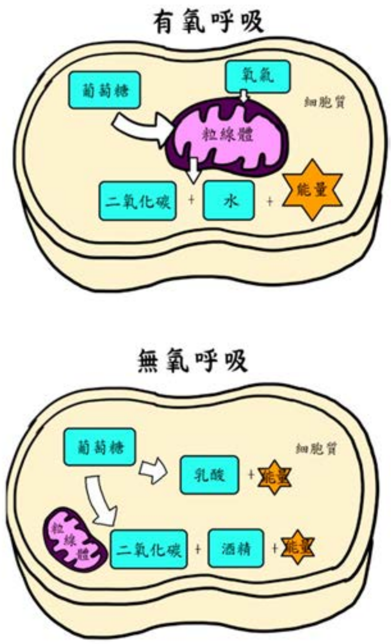

Figure 3. The illustrated passage of cellular respiration.

(1993), the three passages are instructional-level texts for fourth graders and independent-level texts for sixth graders. The same text was used in the two reading conditions. In the text-only group, the text was aligned exactly to the centre of the display. In the illustrated

最近臺東、臺南、花蓮等地都引進熱氣球活動，形成一股搶搭熱氣球的觀光風潮。相傳最早的熱氣球是中國人發明的天燈，當時是為了傳播軍事信號，無法載人。

現代可以載人的熱氣球，由球囊、噴火器及吊籃所組成。噴火器加熱球囊內的空氣，利用空氣熱脹冷縮所產生的浮力作用而升空。

熱氣球在未加熱之前，球囊內、外的空氣溫度相同。當噴火器開始加熱，球囊內的空氣溫度升高。由於球囊內的空氣受熱膨脹，但是球囊本身的容積有限，因此部分的空氣分子會從球囊的底部跑出去。最後，相較於外界的空氣，球囊內的空氣分子較少、重量變輕、密度較小。由於球囊內的空氣密度較小，產生了向上的浮力作用，因此熱氣球就可以升空。

此外，噴火器加熱球囊內部的空氣，除了可使熱氣球升空之外，也可以維持球囊內、外的氣壓相同，讓球囊飽滿。球囊內部的空氣分子受熱後，空氣分子的移動速度加快，因此球囊內部的空氣分子雖然較少，但仍可維持與外界相同的氣壓。

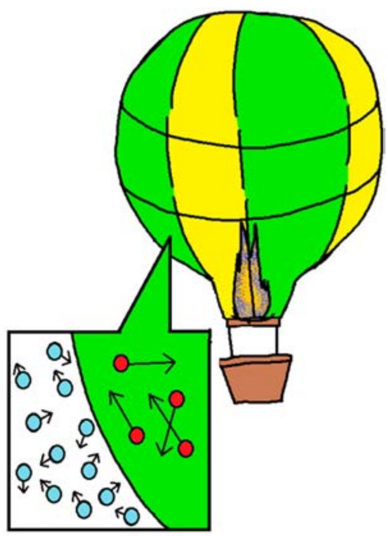

Figure 4. The illustrated passage of hot air balloon.

group, the text was presented on the left side and the other side had the illustration. In addition, the text was all in black and white and the illustrations were all in colour. All illustrations in this study adopted arrow symbols, which were classified into different semantic roles according to Kurata and Egenhofer (2008).

The passage about firefly was 391 Chinese characters long (Figure 2). The subcategory of the arrow symbols' semantic role in this passage was 'labelling', which falls under the major category of 'annotation', indicating the firefly's abdomen in the picture with descriptions (i.e. labels). The first paragraph described the living environment of a firefly. The second and the third paragraphs then mentioned the main components and operating mechanism of a firefly's light-emitting phenomenon. The fourth paragraph explained the critical function of a firefly's reflective cells and used car headlights as a metaphor. Finally, the last paragraph concluded by explaining that a firefly's glow is necessary to their reproduction.

The passage on cellular respiration was 381 Chinese characters long (Figure 3). There were two figures in this passage to represent aerobic and anaerobic respiration. There were two arrow symbol subcategories in the figure depicting aerobic respiration. The first was 'spatial movement', which falls under the major category of 'behavioural description', indicating the movements of glucose and oxygen into mitochondria, which then cause a chemical reaction. The second was 'change', which falls under the major category of 'association', representing the state after chemical reaction (i.e. the combination of carbon dioxide, water, and energy). The figure for anaerobic respiration contained only the subcategory of 'change', which shared identical arrow symbol semantics with the figure for aerobic respiration. The first paragraph introduced cellular respiration by giving the definition. The second and the third paragraphs explained the differences between aerobic and anaerobic respiration. The fourth paragraph provided some concise instances of the two kinds of respiration. Finally, the last paragraph mentioned the impact of aerobic and anaerobic respiration on living creatures.

The passage about hot air balloon was 389 Chinese characters long (Figure 4). The subcategory of the arrow symbols' semantic role in this passage was 'vector', which falls under the major category of 'orientation'. As shown in Figure 4, the tail slots of the arrow symbols were attached to the air molecules, indicating their speed by the length of body slots and the direction with head slots. The first paragraph introduced hot air balloons by providing context and history. The second paragraph described the basic construction of hot air balloons. The third paragraph explained the buoyancy principle of hot air balloons. The final paragraph offered the theory of the operation of hot air balloons and a conclusion.

To assess the students' comprehension of the reading materials, eight questions were developed for each passage on a separate sheet of paper. The test format includes true/false and multiple-choice questions. Two types of questions – factual and comprehension – were included in the comprehension test. Factual questions concerned basic information mentioned in the text, for example: 'The air reaches the light-producing cells of fireflies through their nostrils and tracheoles'. The comprehension questions were about integrating certain information, for example: 'Why is the light of a firefly called bioluminescence? Please choose the correct answer: (1) Because the temperature of a firefly's light is low; (2) Because the colour of a firefly's light is blue; (3) Because a firefly shows up at night when it is the coldest; (4) Because the light of a firefly can freeze the dew.' The three sets of questions have such acceptable internal consistencies that the medium value of the coefficient obtained with Kuder-Richardson formula (20) was about .53. The value was acceptable, for the low ability students were excluded by the reading comprehension screening test, and thus, the differences between individuals were small.

For the pre-test measurement, Ko (1999)'s reading comprehension screening test was conducted to exclude students who may have reading difficulties. Thirty multiple-choice questions composed the test and two dimensions were involved (i.e. local processing and text processing). Its coefficients of internal consistency reliability and test-retest reliability were .80 and .85. Additionally, the scores of the last Chinese/science achievement test taken during the semester were also collected to ensure the equivalence of the two groups created by random assignment.

Apparatus

Eye movements were collected by a Tobii X120 eye tracker at a sampling rate of 120 Hz. The head movement box was about 30 cm × 22 cm × 30 cm, and the distance between participants and monitor was about 70 cm. We presented all learning materials on a 19-inch LCD monitor with 1024 × 768 pixels. The segments of text were the same on the screen, each measuring about 2.5 – 8.5 cm × 19 cm (80 – 245 × 595 pixels), and the segments of the illustration were also identical on the screen, each measuring about 12 – 17 cm × 14 cm (370 – 600 × 430 pixels). The data were recorded with Tobii-Studio (2.2) software.

Procedure

There were three sessions of data collection in this study. First, the reading comprehension screening test was conducted collectively for 143 participants in 25 min. Those participants with reading difficulties were excluded. After obtaining parental consent, students were randomly assigned into two reading groups: illustrated and text-only. The Chinese/science achievement test scores of all participants were also collected.

Second, participants engaged in the eye-tracking experiment individually in a quiet classroom. After instructions, each participant was calibrated with a 5-point procedure in the eye-tracker. Later, the participant was asked to read the practice material carefully, and then press the space bar to finish it. The participant was then given two questions about the practice material on a sheet of paper. The use of practice material was to familiarise our participants with the experimental procedure. The process of formal experiment was identical to the practice one, but the participants were given eight comprehension questions on each passage. The sequence of the three passages for each participant was selected randomly. Participants could complete the materials and the tests at their own pace. The average period of time for a participant to finish reading the material and complete the tests was about 30 min. The comprehension data of 64 students were included in this study, but several participants' eye movements data were excluded (see the Results section). After this session, 44 students participated in the follow-up interview.

Third, the follow-up interviews were conducted individually. First, participants in both groups were asked to read three illustrated passages on screen, which were exactly the same as the testing materials used for the illustrated group. The participants were instructed to think about the semantics of the arrow symbols in each illustration carefully. In addition to the illustrations of firefly and hot air balloon, since the semantics of the arrow symbols in the aerobic and anaerobic respiration figures are similar, only the illustration of aerobic respiration was adopted in the interview session. The instruction is as follows: 'There are several arrows in this illustration. Please try to explain the meaning of those arrows'. Next, the participants had to give their interpretation of these arrow symbols' semantic roles based on the text of the three passages. The sequence of the passages for each participant was the same as in the second session, and the average period of time of each interview was about $5-10$ min.

Data analysis

The reading comprehension screening test comprised 30 items and the possible scores ranged from 0 to 30. We excluded the students with scores lower than 14 since they might have reading difficulties. The reading comprehension test of each passage contained eight questions and their scores would be converted into percentages. Additionally, the scores of the Chinese and science achievement test ranged from 0 to 100.

Based on the time the participant spent on passages, two eye-movement indicators were adopted. First, the fixation duration for the entire passage was calculated by adding the fixation durations of all of its text segments, and the illustrations was also calculated by adding together those of all of the illustration segments. Second, the total fixation duration for each paragraph (milliseconds) was divided by the area of each text paragraph (number of pixels) and referred to as the FD/P ratio in this study. The reason this ratio is calculated is because it makes the paragraphs comparable, while controlling their length. Another analysis of eye movement was the reading sequence, which was conducted only in the illustrated group. It was calculated with sequential analysis (Bakeman & Gottman, 1997), which was conducted via a series of matrix multiplications. According to the paragraphs and illustration(s), the three passages were divided into five to seven areas of interest (AOIs). The fixation transition from each of the AOIs to others was calculated, and the adjusted residuals were analysed using a Z-test to determine whether the observed frequency was significantly greater than expected.

Regarding the participants' interpretations of the arrow symbols that were derived from the interviews, we adopted the classifications proposed by Kurata and Egenhofer (2008) with slight modifications based on the interview results. This is shown in Table 1. The major categories and the subcategories and their scoring criteria were classified based on the keywords and semantic meanings provided by the participants. As mentioned in the 'Materials' section, the arrow symbols of the three passages had two semantic roles each – the major category and subcategory. Take the passage on hot air balloon for example; if a participant stated: 'the arrows represent the moving direction of air molecules', the response would be categorised as the subcategory of 'direction', which is under the major category of 'orientation'. Since the arrow's meaning of the hot air balloon should be 'vector', which falls under the major category of 'orientation', this participant's response would be deemed incorrect for subcategory and correct for major category. Certain data were not available because the semantic meaning could not be identified. The agreement of the two raters was 89% (156/176),

Table 1. The scoring criteria to examine participants' interpretation of eight arrow symbol subcategories.

Major categories and subcategories	Scoring
Orientation	
Direction	Participants refer the words of 'direction' to represent the moving state of property, but no information about the distance nor destination, e.g. the moving direction of molecules.
Vector	Participants mention direction and magnitude simultaneously, e.g. the speed of molecules.
Magnitude	Participants mention distance, destination or quantity, e.g. the degree of atmosphere pressure.
Annotation	
Labelling	Participants use annotations to specify a description of a component, e.g. the position of a structure.
Behavioural description	
Spatial movement	Participants mention the spatial transition from one component to another, e.g. air rushes into a firefly's abdomen through the spiracle, and move into the trachea.
Association	
Conditional relation	Participants refer preconditions prior to the follow-up movements or events, e.g. the air is necessary for decomposing glucose.
Change	Participants mention the transforming of substances, e.g. the production of carbon dioxide, water, and energy.
Ordered relation	Participants refer the levels of order, e.g. the spiracle is a body part of firefly.

and the inter-rater reliability was acceptable, with kappa values of 0.85 ($p < .001$). All disagreements were resolved through discussion.

Based on the participants' responses, the hit rates and the false-alarm rates were calculated. Take 'vector' for example: the hit rate would be the proportion of the participants who correctly recognised 'vector' in the passage of hot air balloon, and the false-alarm rate would be the average proportion of the participants who incorrectly recognised the arrows' meanings illustrated in the passages of firefly and cellular respiration as 'vector'.

In addition, *d'* was calculated by calculating the *z* score of hit rate minus the *z* score of false-alarm rate. The higher the value of *d'*, the more capable the participants are of recognising the semantic roles of arrow symbols. This study defined the acceptance criterion of *d'* as 2. That is, if *d'* for a certain arrow symbol approaches 2, the participants are able to discriminate the correct semantic meanings from the incorrect ones at above-chance levels. For each passage, the hit rate, false-alarm rate, and *d'* would be calculated for both the major category and subcategory of arrows. Except for the firefly passage, since its major category has only one subcategory, the hit rate, false-alarm rate, and *d'* were the same for both major category and subcategory.

Results

Three independent sample *t*-tests were performed to examine whether the different groups had various effects on the pre-test measurements, including the students' reading comprehension screening test, the Chinese achievement test, and the science achievement test (see the top half of Table 2). The results showed no significant difference between the two groups in the three test scores ($ts < 1$, $ps > .10$). This demonstrated that the participants in this study were randomly assigned to the illustrated group or the text-only one.

Comprehension

Three independent sample t-tests were performed to examine whether the comprehension of the three passages were different in the two groups (see the bottom half of Table 2). The two groups showed no significant differences for all passages ($ts < 1$) with a threshold of $p < .01$.

Eye-movement behaviours

The eye movement data contained 64 participants, and six students with the valid ratio below 40% or apparent drift in all three passages were excluded. In addition, the numbers of students whose data were excluded for the three passages (i.e. firefly, cellular respiration, and hot air balloon) were one, three, and one. Therefore, the data were collected from 57 (illustrated/text-only, $n = 28/29$), 55 (illustrated/text-only, $n = 29/26$), and 57 (illustrated/text-only, $n = 28/29$) students, respectively, for inclusion in the analyses. The medians of the valid eye-movement data were .91, .89, and .89, respectively.

An independent sample t-test was performed to examine whether the total fixation duration was different in the two groups (Table 3). With a threshold of $p < .01$, the result showed no significant difference between the two groups in all passages ($ts < 1$). In addition, the fixation durations of the illustration sections of the three passages were all less than ten seconds, and their fixation duration ratio ranged from 5 to 11%.

Table 3 also shows the means and standard deviations for both groups of the four (or five) text paragraphs in terms of the FD/P ratios. We analysed the groups (2) × text paragraphs (4 or 5) with three separate mixed ANOVA design by the FD/P ratios. We found a main effect for the text paragraphs of firefly, $F(4, 220) = 101.65$, $p < .001$, partial $\eta^2 = .65$, cellular respiration, $F(4, 212) = 85.31$, $p < .001$, partial $\eta^2 = .62$, and hot air balloon, $F(3, 165) = 87.56$, $p < .001$, partial $\eta^2 = .61$, but no main effect of the group or interaction effects ($ps > .05$). It indicated that the two groups did not differ in their dwell times even for the structure-related paragraphs (i.e. paragraphs 2, 3, and 4).

For the passage on firefly, post hoc tests (Scheffé) indicated that the FD/P ratios were especially greater for participants in the second, third, and fourth paragraphs ($M = 0.28$, 0.34, and 0.19) than they were in the first and fifth paragraphs ($M = 0.11$ and .08). In the passage on cellular respiration, the FD/P ratio for participants in the third paragraph ($M = 0.37$) was greater than those of the other paragraphs ($M = 0.12$, 0.16, 0.20, and 0.14). Turning to the passage on hot air balloon, the FD/P ratios for participants in

Table 2. Means and standard deviations for pre-test and post-test measurements as a function of group.

	Text-only	Illustrated
Measurements	*M*(*SD*)	*M*(*SD*)
Pre-test measurements		
Reading comprehension screening test	22.31 (4.19)	22.14 (4.21)
Chinese achievement score	91.51 (5.34)	91.98 (6.54)
Science achievement score	78.79 (13.79)	83.68 (12.64)
Post-test measurements		
Comprehension test of firefly	66 (21)	68 (20)
Comprehension test of cellular respiration	67 (18)	67 (21)
Comprehension test of hot air balloon	73 (24)	70 (18)

Table 3. Means and standard deviations for eye movement measures as a function of version.

	Fireflies		Cellular respiration		Hot air balloons	
	Text-only	Illustrated	Text-only	Illustrated	Text-only	Illustrated
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
FD						
Text (s)	68.07 (26.77)	74.13 (34.73)	71.57 (31.74)	76.72 (32.35)	63.77 (28.36)	71.27 (33.49)
Illustration (s)	–	5.50 (5.60)	–	8.95 (7.46)	–	2.98 (3.05)
FD/*P* ratio (ms/pixel)						
First paragraph	0.10(0.06)	0.11(0.05)	0.11(0.06)	0.14(0.07)	0.16(0.08)	0.21(0.08)
Second paragraph	0.29(0.11)	0.28(0.09)	0.15(0.07)	0.17(0.06)	0.18(0.07)	0.18(0.05)
Third paragraph	0.33(0.08)	0.36(0.08)	0.36(0.08)	0.37(0.08)	0.41(0.09)	0.42(0.08)
Fourth paragraph	0.19(0.08)	0.18(0.08)	0.22(0.08)	0.18(0.07)	0.25(0.08)	0.20(0.09)
Fifth paragraph	0.08(0.06)	0.07(0.04)	0.15(0.06)	0.14(0.07)	–	–

the third and fourth paragraphs ($M = 0.41$ and 0.22) were greater than those in the first and second paragraphs ($M = 0.18$ and 0.18).

Regarding the reading sequences, we examined whether the illustrated group demonstrated more text-illustration referencing behaviours on the structure-related paragraphs, that is, paragraph 2 to paragraph 3 (or 4) (Figure 5). The Z-value matrix of subject in three passages are given in the Appendix. Students tended to look back and forth between the paragraphs but seldom refer to illustration(s) among the three passages. Though students could read the illustrations, their referencing behaviours between the structure-related paragraphs and the illustrations were limited. For the firefly passage, the direction of the participants' text-illustration referencing behaviours was illustration toward paragraph 1, with a significantly higher transfer probability than other AOIs ($Z = 3.94$, $p < .001$). For the cellular respiration passage, the direction of the participants' text-illustration referencing behaviours was paragraph 1 toward illustration, with a marginally significant higher transfer probability than other AOIs ($Z = 1.78$, $p = .090$). For the hot air balloon passage, the direction of the participants' text-illustration referencing behaviours was illustration toward paragraph 3, with a significantly higher transfer probability than other AOIs ($Z = 2.\ 42$, $p = .010$). Except for the hot air balloon passage, the participants could not reference the illustration and the structure-related paragraphs, and in appropriate timing.

Participants' interpretation of arrow

We examined young readers' responses in the follow-up interviews to determine whether they could distinguish the arrow symbols' semantic roles. As shown in Table 4, the participants' average hit rate for the major categories of the arrow symbols' semantic roles in the three passages was about .70, though the false-alarm rates contained huge discrepancies (ranging from .01 to .18), with d' ranging from 1.65 to 3.01. 'Annotation' showed a relatively high hit rate and the highest false-alarm rate, indicating that the participants used this meaning frequently and incorrectly. 'Association' and 'orientation' both displayed acceptable hit rates, indicating that most students could aware these two major categories of arrow symbols in illustrated science texts.

The hit rates and the false-alarm rates for the subcategories showed identification discrepancies, and the d' was less than 2 for all (ranging from 1.65 to 1.87), indicating that the participants did not perform well on the subcategories of semantic roles for arrows.

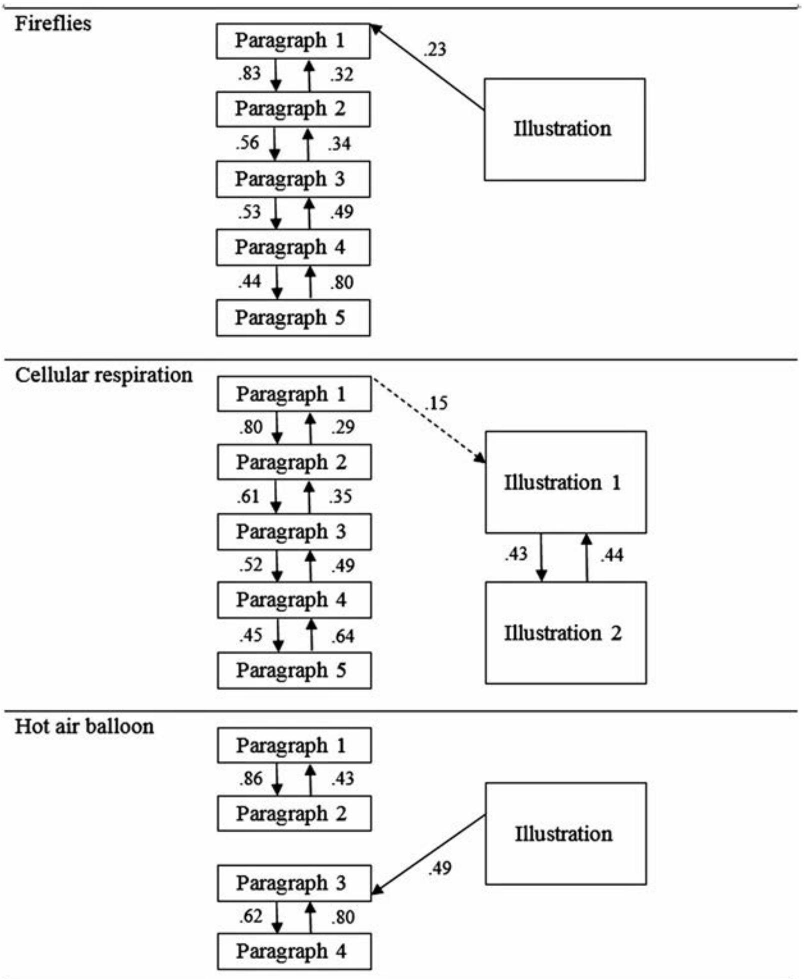

Figure 5. The total-pass transition diagrams in the three passages. Note. The solid arrow in the figure represents significant transition probabilities, and the dotted arrow represents marginally significant ones. The numbers beside the arrow indicate the transition probabilities.

'Labelling' contained the same data as 'annotation', showing that young readers often misuse this semantic role. For 'spatial movement or conditional relation', though the hit rate was only 66%, its false-alarm rate was not that high, indicating that the

Table 4. The hit rate, false alarm rate, and *d′* of distinguishing arrows in participants.

Major and sub-categories	Hit rate	False alarm rate	*d′*
Annotation	.77	.18	1.65
Labelling	.77	.18	1.65
Association	.47	.01	2.45
Spatial movement or Conditional relation	.66	.08	1.84
Change	.57	.05	1.87
Orientation	.80	.02	3.01
Vector	.18	.00	1.75

participants might be slightly better in discriminating the meaning of 'spatial movement or conditional relation' over other subcategories. For 'change', the trends for hit rate and false-alarm rate were similar to those of 'spatial movement or conditional relation', showing that the participants were marginally able to identify the meaning of 'change'. As for 'vector', the hit rate and false-alarm rate were both low, indicating that the participants rarely used this subcategory.

Most participants were inclined to consider 'vector' in an oversimplified way as 'direction' or 'magnitude', revealing that it was difficult for them to aware both 'direction' and 'magnitude' simultaneously. We found that the participants' rate of identifying 'vector' as 'direction' or 'magnitude' was 48% and 14%, respectively. In addition, we also found that the participants frequently confused 'labelling', 'spatial movement', and 'change', probably because of their similar structures – an arrow symbol connecting two components with its head and tail. For example, the rate at which participants mistook 'spatial movement' for 'change' was 14% for the illustration of aerobic respiration.

The most frequently used words to refer to the semantic roles of the arrow symbols

The words most frequently used by young readers to refer to the subcategories of the arrow symbols' semantic roles were also identified. Regarding 'direction', most participants presented components that indicated the process of moving in a certain direction with words such as 'move into', 'flow', and 'moving direction' (e.g. the *moving direction* of molecules and the *flow* of air). For 'vector', participants indicated 'direction' and 'magnitude' simultaneously using words such as 'moving direction with distance/scope' and 'speed' (e.g. the *speed* of molecules). Concerning 'magnitude', participants demonstrated the magnitude of force using words such as 'moving scope', 'pressure', 'magnitude', 'numerous', and 'few' (e.g. the *pressure* of the air outside is low). For 'labelling', participants indicated this effect using words such as 'mean', 'show', 'where', 'what', 'position', 'name', and 'equal' (e.g. the *name* of a specific object). For 'spatial movement', the participants presented the continuous changes of spatial position with words such as 'get into', 'go', 'enter', 'pass through', and 'go through' (e.g. glucose *enters* the mitochondria). Regarding 'conditional relation', participants showed the special relations between two components with words like 'it has to do something' and 'it needs something' (e.g. mitochondria *need* glucose and oxygen). For 'change', participants presented the changes between two components by words such as 'become', 'generate', and 'turn into' (e.g. something *turns into* carbon dioxide, water, and energy). Concerning 'ordered relation', participants presented the order between two components using words such as 'under', 'include', and 'A is part of B' (e.g. oxygen, glucose, and energy are all *included* in the mitochondria).

Discussion

This study found no evidences that young readers could utilise scientific illustrations. For Hypothesis 1, there was no significant difference between the illustrated and text-only groups in terms of comprehension. For Hypothesis 2, we found no significant difference for the text section's total fixation duration between the groups. Moreover, the fixation durations of the illustration sections for the groups were less than ten seconds. For Hypothesis 3, the analyses of FD/P ratios showed no group effect in terms of the paragraphs with structure-related information (i.e. paragraphs 2, 3, and 4). However, the main effect of the text paragraphs revealed that students in both groups could identify the importance of the second and third paragraphs. For Hypothesis 4, the illustrated group demonstrated limited referencing behaviours between text and illustrations, and had inappropriate timing. In general, the two groups had similar reading comprehension with no ceiling effect. The groups took the same time for the whole text and for the structure-related paragraphs. These results revealed that young readers do not utilise the information from scientific illustrations. This explained why there was no significant difference in comprehension and eye-movement behaviours between the two groups.

These findings are open to two alternative explanations. First, the test items might contain little information that could be acquired only from the illustrations, thus showed no advantages in the illustrations. Second, the texts might be less ambiguous to the readers for its structure-related information. Thus, the positive effects of illustrations, for example, constraining interpretations and increasing comprehension, would be less pronounced. Both explanations could only explain the insignificant comprehension performances between the two groups, but not about why the illustrated group's reading time on the text sections and particularly the structure-related paragraphs were not shorter than the text-only group. Even if the text was less ambiguous, the picture could accelerate the processing of structure-related paragraphs because part of the mental model was constructed by inspecting the picture (Eitel et al., 2013). Hence, the only reasonable explanation is that young readers rarely inspected illustrations. Young readers' reading sequences provides further evidence that their referencing behaviours between structure-related paragraphs and illustration were limited, and had inappropriate timing. It may be that students simply could not connect the illustrations with the text or did not believe that specific illustrations could also be a source of knowledge, and thus only spent a few seconds on them. Another possible explanation for young readers' limited inspecting time on illustrations is that our texts already had all the information; thus, there was no need to acquire extra information from the illustrations. The finding corresponds to the previous research that young readers' reading time of illustrations were short and the referencing behaviour between texts and illustrations was also insufficient (Hannus & Hyönä, 1999; Jian, 2016; Jian & Ko, 2017). Besides, this study revealed that the scaffolding assumption of Eitel et al. (2013) was not supported. We suggested that briefly inspecting the illustrations might not be enough for young readers to construct a partial mental model relating to the structure. Thus, their comprehension and text processing could not be improved. In this study, we adopted only close-ended questions to examine participants' reading comprehension, which might not be adequate to elicit the benefits of illustrations. We believe that more open-ended questions would be needed to examine whether a brief consideration of the illustration is enough for young readers to form a deeper conceptual understanding and develop mental models.

For the exploratory research questions 1 and 2, we found that the participants could identify most of the major categories with their words. However, they encountered more difficulties while recognising the semantic roles of subcategories. Their performances on 'labelling' contained the highest false-alarm rate, demonstrating its overextension. This is to say that the young readers were inclined to refer to their familiar semantic roles (i.e.

labelling) by using words like 'mean', 'show', 'where', 'what', 'position', 'name', and 'equal' to decipher every arrow in scientific diagrams. In addition, the participants' performances on the 'vector' contained the lowest hit rate, indicating an underextend use of this meaning. The young readers failed to appropriately mention the semantic roles of 'vector' with words such as 'moving direction with distance/scope' and 'speed'. This might be because 'vector' is a compound concept, which comprises of both 'direction' and 'magnitude'. Most of the young readers might not have developed this concept; thus, they tended to oversimplify its meaning. The concept of overextension and underextension originated from the theory of language development (e.g. Hoff, 2013; Kay & Anglin, 1982) and is related to Jean Piaget's concept of 'assimilation' (Piaget & Cook, 1952), which alludes to the idea that individuals tend to use existing schema to cope with new situations. Besides, some participants confused 'labelling', 'spatial movement', and 'change', which is probably because similar structures of these arrow symbols were difficult for young readers to distinguish.

As discussed above, we found that young readers could employ various semantic roles of arrow symbols with their own words, even though they might not have been explicitly taught about it. This corresponds to the studies of language development that posit a large proportion of word/vocabulary growth to occur through incidental learning, that is, from normal reading (Nagy et al., 1987; Rott, 1999). Rott (1999)'s empirical study found that learners' vocabularies increased as long as they encountered the unfamiliar words twice. Similarly, the young readers should be able to learn various semantic roles of arrow symbols from written context during their school years. Nonetheless, the young readers did not perform very well on the recognition performances, especially for the subcategories of arrows' meanings, indicating the difficulty in understanding unfamiliar arrows' meanings and new scientific knowledge simultaneously. Thus, it should be necessary to assist young readers in understanding and distinguishing the critical conventional representations in diagrams (i.e. arrow symbols), as suggested by Cromley, Bergey, et al. (2013) and Cromley, Perez, et al. (2013).

Due to the inadequacy of empirical study in classifying the meanings of the arrow symbols, we analyse our interview data with the structure of Kurata and Egenhofer (2008). The structure of arrows' semantic roles was identified by a computer algorithm. We found that our young readers' interpretations can indeed be classified into Kurata and Egenhofer's structure. This indicates that the psychological reality of Kurata and Egenhofer's semantic structure was confirmed by our empirical evidences.

Implications and future studies

This study has several pedagogical implications. First, many students may have already possessed adequate reading comprehension strategies for texts but lack proper strategies for reading illustrations. Thus, teachers should help students to understand the connections between critical paragraphs and illustrations and then strengthen their strategies for integrating text and illustrations, such as adopting indirect instruction tools to improve purposeful processing (Mason et al., 2017) or signals to make the related information in pictures noticeable (Ozcelik et al., 2009).

Second, students tended to overextend and underextend certain semantic roles of arrow symbols. Thus, teachers should make it a point to help learners to distinguish the meaning of ‘annotation’ from others to decrease its overextended application and make them aware of the meaning of ‘vector’ to avoid its underextended application. Furthermore, students occasionally confused some semantic roles of arrow symbols, such as the semantic roles of ‘labelling’, ‘spatial movement’, and ‘change’. To resolve this confusion, teachers should assist students in distinguishing the structures and meanings of the arrow symbols’ semantic roles. For instance, if there is an arrow connecting a description and a subject with its head and tail, it should be described as ‘labelling’ (e.g. an arrow that connects the description of a firefly’s neuron and the subject); if there is an arrow indicating different states of a subject with its head and tail, it should be ‘change’ (e.g. an arrow demonstrating the transformation of glucose and oxygen into carbon dioxide, water, and energy); if there is an arrow presenting the movement of a subject with its head and tail, it should be ‘spatial movement’ (e.g. an arrow indicating the spatial movement of glucose into mitochondria).

Several directions for future research in illustrated science texts could be considered. First, this study included only participants with normal reading ability, and they were not sorted into groups by ability. For future research, researchers may consider investigating the difference in reading strategies or performances between poor, average, and skilled comprehenders. Second, the learning materials selected were three biology and physics illustrated passages, which were classified as Clay (1993)’s independent reading level. This meant that the participants could comprehend most of the content without instruction. Nonetheless, more illustrated science passages in various domains or reading levels should be adopted to investigate the generalisability of the finding. Third, one of the possible reasons why young students paid little attention to the illustrations is because the texts were understandable without the presence of the illustrations. Future studies should confirm if the illustration is more informative than the text before one selects reading materials, and further compare their effects on young readers’ reading behaviours. Fourth, this study found that young readers tend to focus on the text section and demonstrate text-directed strategy. However, one might argue that the position of our learning materials–the text on the left and the illustration on the right–had a significant influence on young readers’ reading behaviours. Future researchers should conduct an additional study by exchanging the sides of text and illustration to exclude this alternative explanation. Fifth, we adopted the coding structure of Kurata and Egenhofer (2008) to analyse the participants’ interpretation of arrows. However, there are still other meanings of arrows to be used in real diagrams. For instance, the arrows could be used to demonstrate a transformation over time, possible sequences, or enlargements. Future research should further develop and confirm the initial coding structure of arrows. Sixth, we identified the words that young readers used most frequently to refer to the arrow symbols. More studies are needed to examine whether adopting these words in instruction would help students to understand various functions of arrows, or would instead distract, constrain, or mislead readers’ interpretations of diagrams if they focus too much on the arrows.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Ministry of Education [grant number Higher Education Sprout Project]; Ministry of Science and Technology, Taiwan [grant number MOST 108-2410-H-003-046-MY2, NSC 101-2511-S-003-015].

ORCID

Chao-Jung Wu http://orcid.org/0000-0001-7156-1487
Chia-Yu Liu http://orcid.org/0000-0002-2837-3461

References

Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, *125*, 413–428. https://doi.org/10.1016/j.compedu.2018.06.023

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

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 & Instruction*, *20*(2), 123–135. https://doi.org/10.1016/j.learninstruc.2009.02.015

Bowen, G. M., & Roth, W.-M. (2002). Why students may not learn to interpret scientific inscriptions. *Research in Science Education*, *32*(3), 303–327. https://doi.org/10.1023/A:1020833231966

Clay, M. M. (1993). *An observation survey of early literacy achievement*. Heinemann.

Coleman, J. M., McTigue, E. M., & Smolkin, L. B. (2011). Elementary teachers' use of graphical representations in science teaching. *Journal of Science Teacher Education*, *22*(7), 613–643. https://doi.org/10.1007/s10972-010-9204-1

Crocker, N., Vaurio, L., Riley, E. P., & Mattson, S. N. (2011). Comparison of verbal learning and memory in children with heavy prenatal alcohol exposure or attention-deficit/hyperactivity disorder. *Alcoholism: Clinical and Experimental Research*, *35*(6), 1114–1121. https://doi.org/10.1111/j.1530-0277.2011.01444.x

Cromley, J. G., Bergey, B. W., Fitzhugh, S. L., Newcombe, N., Wills, T. W., Shipley, T. F., & Tanaka, J. C. (2013). Effects of three diagram instruction methods on transfer of diagram comprehension skills: The critical role of inference while learning. *Learning and Instruction*, *26*(1), 45–58. https://doi.org/10.1016/j.learninstruc.2013.01.003

Cromley, J. G., Perez, A. C., Fitzhugh, S., Newcombe, N., Wills, T. W., & Tanaka, J. C. (2013). Improving students' diagram comprehension with classroom instruction. *Journal of Experimental Education*, *81*(4), 511–537. https://doi.org/10.1080/00220973.2012.745465

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

Elen, J., & Gorp, E. V. (2008). The effects of multimedia design features on primary school learning materials. *International Journal of Instructional Media*, *35*(1), 7–15.

Hannus, M., & Hyönä, J. (1999). Utilization of illustrations during learning of science textbook passages among low-and high-ability children. *Contemporary Educational Psychology*, *24*(2), 95–123. https://doi.org/10.1006/ceps.1998.0987

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

Hoff, E. (2013). *Language development*. Wadsworth.

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. https://doi.org/10.1002/rrq.125

Jian, Y. C. (2020). Teaching fourth-grade students of different reading abilities to read biological illustrations and integrate in-text information: An empirical experiment. *Research in Science Education*, *50*, 2269–2282. https://doi.org/10.1007/s11165-018-9778-8

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. , & 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.

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*, 51–62. http://dx.doi.org/10.1016/j.learninstruc.2014.01.005

Kay, D. A., & Anglin, J. M. (1982). Overextension and underextension in the child's expressive and receptive speech. *Journal of Child Language*, *9*(1), 83–98. https://doi.org/10.1017/S0305000900003639

Ko, H. W. (1999). Reading comprehension-screening test [in Chinese]. *Psychological Testing*, *46*, 1–11.

Kramer, J. H., Rosen, H. J., Du, A. T., Schuff, N., Hollnagel, C., Weiner, M. W., & Miller, B. L. (2005). Dissociations in hippocampal and frontal contributions to episodic memory performance. *Neuropsychology*, *19*(6), 799–805. https://doi.org/10.1037/0894-4105.19.6.799

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

Kurata, Y., & Egenhofer, M. (2008). The arrow-semantics interpreter. *Spatial Cognition and Computations*, *8*(4), 306–332. https://doi.org/10.1080/13875860802148843

Lee, V. R. (2010). Adaptations and continuities in the use and design of visual representations in US middle school science textbooks. *International Journal of Science Education*, *32*(8), 1099–1126. https://doi.org/10.1080/09500690903253916

Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Communication and Technology*, *30*, 195–232. https://doi.org/10.1007/BF02765184

Levin, J. R., Anglin, G. J., & Carney, R. N. (1987). On empirically validating functions of pictures in prose. In D. M. Willows, & H. A. Houghton (Eds.), *The psychology of illustration, Vol. 1: Basic research* (pp. 51–85). Springer.

Liu, Y., & Khine, M. S. (2016). Content analysis of the diagrammatic representations of primary science textbooks. *Eurasia Journal of Mathematics, Science and Technology Education*, *12*(8), 1937–1951. https://doi.org/10.12973/eurasia.2016.1288a

Mason, L., Pluchino, P., & Tornatora, M. C. (2013). Effects of picture labeling on science text processing and learning: Evidence from eye movements. *Reading Research Quarterly*, *48*(2), 199–214. https://doi.org/10.1002/rrq.41

Mason, L., Scheitert, K., & Tornatora, M. C. (2017). Using eye movements to model the sequence of text–picture processing for multimedia comprehension. *Journal of Computer Assisted Learning*, *33*(5), 443–460. https://doi.org/10.1111/jcal.12191

Mason, L., Tornatora, M. C., & Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers and Education*, *60*(1), 95–109. https://doi.org/10.1016/j.compedu.2012.07.011

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

McTigue, E. M. (2009). Does multimedia learning theory extend to middle-school students? *Contemporary Educational Psychology*, *34*(2), 143–153. https://doi.org/10.1016/j.cedpsych.2008.12.003

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, *24*(2), 237–270. https://doi.org/10.3102/00028312024002237

Ozcelik, E., Karakus, T., Kursun, E., & Cagiltay, K. (2009). An eye-tracking study of how colour coding affects multimedia learning. *Computers and Education*, *53*(2), 445–453. https://doi.org/10.1016/j.compedu.2009.03.002

Piaget, J., & Cook, M. T. (1952). *The origins of intelligence in children*. International University Press.
Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied*, *7*(3), 219–226. https://doi.org/10.1037/1076-98X.7.3.219
Reid, D. J., & Beveridge, M. (1986). Effects of text illustration on children's learning of a school science topic. *British Journal of Educational Psychology*, *56*(3), 294–303. https://doi.org/10.1111/j.2044-8279.1986.tb03042.x
Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, *21*(4), 589–619. https://doi.org/10.1017/S0272263199004039
Russo, M. J., Cohen, G., Campos, J., Martin, M. E., Clarens, M. F., Sabe, L., … Allegri, R. F. (2016). Usefulness of discriminability and response bias indices for the evaluation of recognition memory in mild cognitive impairment and Alzheimer disease. *Dementia and Geriatric Cognitive Disorders*, *19*(6), 799–805. https://doi.org/10.1037/0894-4105.19.6.799
Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 72–103). Cambridge University Press.
Schweppe, J., Eitel, A., & Rummer, R. (2015). The multimedia effect and its stability over time. *Learning and Instruction*, *38*, 24–33. https://doi.org/10.1016/j.learninstruc.2015.03.001
Smith, J. M., & Pol, K. (2018). What's going on in a graphic novel? *Reading Teacher*, *72*(2), 169–177. https://doi.org/10.1002/trtr.1695
Sung, Y. T., Chen, J. L., Cha, J. H., Tseng, H. C., Chang, T. H., & Chang, K. E. (2015). Constructing and validating readability models: Integrating multilevel linguistic features with machine learning methods. *Behaviour Research Methods*, *47*(2), 340–354. https://doi.org/10.3758/s13428-014-0459-x
Tversky, B. (2001). Spatial schemas in depictions. In M. Gattis (Ed.), *Spatial schemas and abstract thought* (pp. 79–111). MIT Press.

Appendix

Z-value matrix of subject in three passages.

Firefly

Target area	Paragraph 1	Paragraph 2	Paragraph 3	Paragraph 4	Paragraph 5	Illustration
Start area						
Paragraph 1		15.54*	−6.16	−5.78	−4.16	0.55
Paragraph 2	10.16*		7.41*	−9.60	−6.48	0.12
Paragraph 3	−5.30	2.50*		7.57*	−7.86	−1.17
Paragraph 4	−6.54	−9.94	4.83*		14.89*	−3.37
Paragraph 5	−3.47	−5.31	−5.91	12.82*		1.10
Illustration	3.94*	0.74	1.32	−2.89	−2.36	

Cellular respiration

Target area	Paragraph 1	Paragraph 2	Paragraph 3	Paragraph 4	Paragraph 5	Illustration1	Illustration 2
Start area							
Paragraph 1		14.83*	−5.28	−5.55	−3.67	1.78	−2.68
Paragraph 2	10.24*		9.43*	−8.77	−5.80	−1.47	−4.40
Paragraph 3	−4.18	4.29*		9.04*	−7.40	−2.81	−4.92
Paragraph 4	−5.59	−8.89	5.42*		15.14*	−5.18	−2.72
Paragraph 5	−3.42	−4.68	−5.72	10.23*		1.21	2.91
Illustration 1	1.39	−0.56	−1.27	−4.93	−2.91		14.22*
Illustration 2	−0.89	−2.88	−1.85	−0.83	−0.33	9.24*	

Hot air balloon

Target area	Paragraph 1	Paragraph 2	Paragraph 3	Paragraph 4	Illustration
Start area					
Paragraph 1		14.17*	−7.61	−6.38	−0.64
Paragraph 2	9.01*		1.27	−8.55	−0.85
Paragraph 3	−6.01	−4.36		11.90*	−2.54
Paragraph 4	−4.42	−7.58	10.03*		0.68
Illustration	−0.99	0.71	2.42*	−2.54	

Note. The rows are the starting areas and the columns are the target areas. *$p < .05$.