STEM Education

STEM Education, 5(3): 448–472
DOI: 10.3934/steme.2025022
Received: 12 February 2025
Revised: 30 March 2025
Accepted: 02 April 2025
Published: 07 May 2025

https://www.aimspress.com/journal/steme

Research article

Visual attention to different types of graphical representations in elementary school mathematics textbooks: An eye-movement-based study

Shumeng Ni[1], Zhujun Jiang[2] and Fengkuang Chiang[2,]*

[1] Department of Linguistics, Shanghai Jiao Tong University, No. 800 Dongchuan Road, Minhang District, Shanghai, China; colormeblue@sjtu.edu.cn
[2] School of Education, Shanghai Jiao Tong University, No. 800 Dongchuan Road, Minhang District, Shanghai, China; zhujunjiang@sjtu.edu.cn, fkchiang@sjtu.edu.cn

* **Correspondence:** Email: fkchiang@sjtu.edu.cn.

Academic Editor: Kwesi Yaro

Abstract: Graphical representations play a crucial role in elementary school mathematics textbooks by conveying information visually and fulfilling specific educational functions. This study examined how seven students aged 9–11 years from Shanghai engaged with graphical elements in an elementary school mathematics textbook. Illustrations were classified into three types: decorative diagrams, guiding diagrams, and informational diagrams. By employing a combination of eye-tracking technology and retrospective interviews, the study analyzed and compared the attention allocation differences among these illustration types. The findings indicated that participants relied more on reading textual content than on illustrations. Moreover, they tended to pay less attention to guiding illustrations and more frequently referred to informational illustrations when reading text. This study identified several limitations in the design of mathematics textbooks and offered practical recommendations for improvement.

Keywords: eye-tracking movement, mathematics textbooks, graphical representations, visual attention patterns, improvement suggestions

1. Introduction

1.1. Background of the study

Information processing theory defines representation as the manner in which information is presented in the mind [1]. The representation process has two primary stages: the acquisition and processing of information. Information is acquired and then undergoes a series of cognitive operations, such as input, encoding, and conversion, before ultimately being represented in the mind. Graphical representations convey information through graphical formats, such as schematics and function graphs, offering a more intuitive and accessible alternative to textual representations [2].

The use of multiple representations is fundamental in educational textbooks. Numerous studies have suggested that learners achieve a deeper understanding of concepts when graphical elements are presented in conjunction with other forms of representation because different visual formats emphasize distinct aspects of information [3,4], particularly when readers can effectively organize and integrate textual and graphical elements on the basis of their prior knowledge [5]. According to dual coding theory, activating both verbal and visual representations in working memory facilitates a connection between the two codes, thereby enhancing information retrieval [6]. However, the extent to which students prioritize textual versus graphical content remains unclear. Most studies suggest that textual content dominates attention due to its direct informational value [3–6,7], while others argue that well-designed graphical elements can significantly enhance comprehension and engagement [8,9].

In mathematics textbooks, graphical representations are commonly referred to as illustrations. Illustrations can be categorized by various criteria, leading to diverse interpretations. Many researchers have classified graphical representations by their functions. Diagrams may be decorative, representational, organizational, explanatory, or transformational [10]. Decorative diagrams contribute little to the comprehension of the accompanying text and are mainly used to add an emotional element. Representational diagrams (also known as analytical diagrams [11]) are closely related to the text they support and accurately depict the concepts described. Organizational diagrams are graphical representations that organize information and components to provide cohesion across chapters. Explanatory diagrams are those that extend beyond organizational diagrams by providing additional details necessary to convey complex or unfamiliar concepts, objects, or phenomena [10]. Transformational diagrams (also known as metaphorical diagrams [12]) refer to graphics that attempt to recode information into a more memorable format [10].

In most cases, researchers employ similar taxonomies but different terminology to describe similar concepts. This problem can be addressed through a careful comparison and integration of existing classification systems. In the present study, the functions of diagrams in mathematics textbooks exhibit subject-specific characteristics. For example, the cartoon character-guided illustrations commonly found in Chinese elementary school mathematics textbooks serve both the organizational function of chapter transitions and the cognitive heuristic role through personified dialogues. Such composite diagrams are classified as “guiding diagrams” in this study. On the other hand, diagrams containing formula derivations and geometric representations, due to their precise visual representation of mathematical concepts, are defined as “informational diagrams”. This classification system inherits Levin’s functional principles while accommodating the diagram characteristics of the mathematical discipline. Therefore, we classify the diagrams in Chinese primary school textbooks into three categories: decorative diagrams, guiding diagrams, and informational diagrams.

While the functional taxonomy of diagrams provides a theoretical lens for analyzing visual representations, their actual utilization in learning contexts may reveal cross-cultural variations. Students' reading habits are not only shaped by learning styles but also by the epistemic climate of classrooms — that is, how educational systems shape students' expectations about knowledge acquisition [13]. Although limited studies have directly compared how Chinese and Western children allocate attention between text and images in textbooks, differences in their classroom epistemic climates suggest potential variations. Western children's object-focused visual processing enhances their ability to extract key graphical information, while Eastern children's context-focused mode relies more on textual cues [14]. As a result, Western children may naturally devote more attention to illustrations, diagrams, and other non-text elements while reading, actively seeking connections between visuals and written content. In contrast, Chinese educational traditions prioritize the close reading of text, with classroom instruction often centered on the teacher's verbal explanations. Images typically serve as supplementary aids rather than independent knowledge carriers, which may lead children to focus primarily on textual content and engage with visuals in a more functional manner—such as quickly locating key diagrams — rather than analyzing them in depth.

1.2. Eye-movement metrics studies and textbook research

The relationship between eye movement indicators and attention mechanisms is grounded in the eye-mind theory proposed by Just and Carpenter (1980) [15]. This theory posits that eye-movement behavior is directly synchronized with cognitive processing, offering a robust framework for interpreting how eye-tracking data reflects attentional dynamics. The theory is built on two foundational assumptions: the immediacy assumption and the eye-mind assumption.

The immediacy assumption states that readers process each word immediately during reading, without significant delay. This explains why gaze duration can serve as a reliable indicator of the depth and difficulty of cognitive processing. The eye-mind assumption suggests that the position of eye fixation directly mirrors the content being processed by the brain, meaning that where the eyes look is where the mind is actively engaged.

Empirical studies supporting this theory demonstrate that participants tend to fixate longer on areas requiring more extensive cognitive processing, such as complex concepts or unfamiliar terms. This alignment between eye movements and cognitive activity provides a strong basis for using eye-tracking metrics to infer attentional mechanisms [16–18].

Building on this theoretical foundation, the relationship between specific eye-movement indicators and attention mechanisms can be articulated as follows:

Total/Mean Gaze Duration reflects sustained attention or deep information processing. A longer total or mean gaze duration typically indicates that the learner has engaged in in-depth cognitive processing, often involving complex or challenging material [19,20]. The pattern between total gaze duration and mean gaze duration may not be consistent. For example, if a graph has a higher total gaze duration but a lower mean gaze duration, it could indicate that participants had multiple fixations or frequent regressions on that graph.

The number of gaze points refers to the total count of gaze points in a designated area of interest (AOI). It reflects the distribution of attention across the material. A higher number of gaze points suggests more frequent shifts in attention, which may indicate either a broader exploration of the material or difficulty in focusing on specific areas [19,21].

Time to first gaze reflects the speed of the initial attention capture. A shorter time to first gaze indicates that the material or specific area is highly salient or engaging, quickly attracting the learner's attention [19,21].

Total/mean visit time refers to the time spent in an AOI, starting from the first gaze point until the participant's gaze moves out of the AOI. They reflect the overall attention allocated to a specific area or material. A longer total or mean visit time suggests sustained engagement and deeper processing, indicating that the area is of significant importance or complexity to the learner [21]. The differences between these two metrics are similar to those previously discussed regarding mean/total gaze duration.

The number of visits refers to the total count of visits to a designated AOI. It reflects the frequency of attention allocation to a specific area. A higher number of visits suggests repeated engagement, which may indicate either the importance of the area or the need for multiple attempts to fully process the information [19,22].

This study will utilize the aforementioned seven metrics for analysis (see 2.4). Overall, “gaze” refers to the duration a learner’s eyes remain in a specific area, typically measured in terms of fixations. A fixation is defined as the eye staying in one position for a certain period (usually 100–200 milliseconds), reflecting attention allocation. “Visit” refers to the grouping of multiple fixations by a learner within the same area [22]. A visit can include multiple fixations, as long as they occur within the same region, indicating repeated focus or attention. These metrics cover multiple aspects of students’ visual attention, including attention allocation, duration, frequency, and early attention. By comprehensively analyzing these metrics, researchers can gain a more holistic understanding of students’ reading behaviors.

1.3. Eye-movement studies on textbook research

Research on graphic reading has largely employed the think-aloud protocol, in which participants verbally articulate their thought processes while solving problems. However, according to cognitive load theory, this approach may interfere with the reading process, particularly for young readers, because it requires them to simultaneously read and verbalize their thoughts, potentially leading to cognitive resource competition [23]. Eye-tracking methodologies offer distinct advantages in graphic reading research. Unlike think-aloud protocols, eye-tracking allows for the real-time recording of reading behavior without imposing additional cognitive load. This method provides detailed quantitative data, including gaze points, gaze duration, and scanning paths. Additionally, eye-tracking visually captures how readers integrate textual and graphical information [16,23].

One of the earliest studies on eye movements in young children while reading illustrations in science texts was conducted by Hannus and Hyona (1999) [17]. This study examined how fourth-grade students with varying cognitive abilities processed illustrations in a biology textbook. The reading material included illustrations with varying degrees of relevance to the accompanying text. For example, in a passage about flies, an illustration depicting the metamorphosis of a fly from egg to adult was considered the most relevant, whereas a photograph of a mosquito was deemed the least relevant. After reading each passage, the students were asked to summarize the main points and answer factual and comprehension questions based on both the text and illustrations. The analyzed

AOIs were the text, illustrations, graphic captions, and blank spaces within each passage. The findings revealed that students with high cognitive abilities outperformed those with low cognitive abilities on all reading tests. Although the total time spent reading illustrations and graphic captions did not considerably differ between the two groups, the students with low cognitive abilities spent more time fixating on blank sections. Further analysis revealed that both the groups of students allocated more time to highly relevant illustrations than to less relevant ones. However, the students with high cognitive ability demonstrated a greater tendency to establish connections between text sections to corresponding illustrations, suggesting that they employed more sophisticated reading strategies focused on synthesizing relevant information.

Studies have experimented with eye-tracking methodologies to investigate reading behavior in relation to graphical representations. One notable finding is that children allocate only approximately 6% of their total reading time to illustrations. By contrast, adults allocate 20%–30%. The ability to distinguish between relevant and irrelevant visual information has been identified as a crucial factor in reading comprehension of illustrated texts [18]. Expanding on their research, Mason et al. (2015) further demonstrated that the purposeful allocation of visual attention is essential for the successful integration of textual and graphical information [24]. In this context, informational charts, which directly support the text, provided stronger pedagogical support than did decorative illustrations, which, although they seemed to elicit emotional responses, had no measurable positive or negative impact on learning [25]. Additionally, although fourth-grade students frequently shifted their gaze between text passages and illustrations, they made relatively few eye movements specifically aimed at cross-referencing the two sources of information. High-achieving students, for example, are more likely to integrate multiple sources of information effectively [26].

1.4. Research hypotheses

Eye-tracking methods have been applied to study the perception of graphical representations in mathematics textbooks in multiple studies, although few of the studies involved categorization of the graphical representations [25]. The present study defined a classification framework and investigated different responses of students to various types of graphical representations, extending beyond a general comparison between graphical and textual content to include distinctions among different types of graphical representations.

Students in grades three through six (aged 9–11 years) are in the stage of transitioning from reading to learning and are engaging with increasingly complex texts [27]. This developmental pattern motivated our focus on this age group to investigate how they allocate attention between textual and graphical representations in mathematics textbooks.

Based on the previous analysis, the following hypotheses are proposed:

Hypothesis 1 (H1): Primary school students will allocate more visual attention to textual content than to graphical content in mathematics textbooks. This hypothesis is grounded in the notion that text serves as the primary source of information, and students may prioritize it over graphical elements [17,24,26].

Hypothesis 2 (H2): Informational graphics will attract more visual attention than decorative or guiding graphics in mathematics textbooks. This hypothesis is supported by findings that informational graphics directly complement the text and are more likely to enhance comprehension [28,29].

Hypothesis 3 (H3): Students who have already grasped the knowledge are more likely to cross-reference between textual and graphical content in mathematics textbooks. This hypothesis is based on evidence that these students possess stronger information integration skills, enabling them to effectively utilize multiple sources of information [29].

The main purpose of this study is to investigate whether these three hypotheses are correct, while also exploring students' attitudes toward textbooks and their representations, and providing suggestions for textbook improvement.

2. Research design

2.1. Participants and measurement instruments

The study included elementary school students from grades 3 to 5. Participants were required to have normal vision or myopia with a degree of 500 degrees or less and had to be available on weekends. Recruitment was conducted through social media, where details of the experiment were posted, and questionnaires were distributed. The study protocol was approved by the Human Research Ethics Committee of the researchers' university (approval number: LL2024000206). Prior to participation, informed consent forms were provided to all students and their guardians, outlining the study's purpose, potential risks, data confidentiality measures, and other relevant information. Both the students and their guardians were fully informed about the study procedures in advance.

A total of seven students (five boys and two girls) aged 9–11 years (mean age, 10 years) from Shanghai were recruited. All participants successfully completed the study. Each student was assigned an identification number in the order of their participation: S1, S2, S3, S4, S5, S6, and S7. Among these, S3 and S5 were girls, while the remaining participants were boys. The profile of the students can be found in Table 1. All the collected data were kept strictly confidential.

Before commencing the eye-tracking experiment, each participant completed a questionnaire adapted from Li Wang's (2018) master's thesis [30]. The questionnaire was divided into two sections. The first section comprised a reading habits survey, consisting of seven items scored on a 5-point Likert scale. The second section comprised three mathematics questions directly related to the study materials, each with a correct answer. They were designed with gradually increasing difficulty levels to assess whether participants had been exposed to the topics and, if so, their mastery of concepts across varying levels of complexity.

The reading habits survey primarily assessed within-group differences. The mean score across the seven items was 1.9184, with a standard deviation of 0.24. The difference between the highest and lowest individual scores was less than 1.00, indicating minimal variation among the participants in terms of their overall attitudes toward and interest in mathematics and their comprehension of the textbook illustrations. Therefore, the data from all seven participants were retained as valid.

The second section questionnaire assessed students' prior mathematical knowledge. The results revealed that the two 11-year-old participants had learned the topic in class and were able to solve problems across varying levels of difficulty (although one 11-year-old participant did not answer the third question correctly, which was due to a careless mistake in unit conversion, and overall, his understanding of the material was not significantly impacted), and that the remaining five participants had not yet studied the material.

Table 1. Participant characteristics.

Participant	Sex	Grade	Age	Whether or not the examined knowledge has been acquired
S1	male	6	11	Yes
S2	male	4	10	No
S3	female	3	9	No
S4	male	3	10	No
S5	female	4	10	No
S6	male	5	11	Yes
S7	male	3	9	No

2.2. Reading materials

The reading material selected for this study was drawn from the "Rectangular and Square" chapter of the fifth-grade mathematics textbook published by the Shanghai Education Edition, specifically the "Volume and Capacity" subsection. This material spans four pages, corresponding to pages 64, 66, 67, and 68 in the original textbook, with page 65 excluded because it solely contained no explanation of new knowledge. The subsection was selected as it exemplifies a typical geometry chapter, particularly within the domain of three-dimensional geometry, which inherently involves a diverse array of graphical representations. After a thorough examination of the entire textbook, it was determined that this chapter contains the highest density and variety of diagrams, making it an optimal choice for this study. The selected pages introduce key concepts related to volume and capacity, unit conversions, volume–mass conversions, and the formula for calculating the volume of a cube.

Building on the classification of graphical representations in textbooks (see Section 1.1), the graphical elements in the selected reading materials were categorized into three primary types on the basis of their characteristics: decorative diagrams, guiding diagrams, and informational diagrams. Decorative diagrams are those that do not substantively support the text but serve to visually represent the graphics mentioned in the text. Guiding diagrams are primarily composed of cartoon characters (e.g., pictures of little boys, little girls, and little pandas), used to trigger verbal expressions or dialogues within the narrative. Informational diagrams are those that directly complement the text and provide accurate visual representations of the text contents. These diagrams are typically placed next to examples or formulas and offer mathematical information or visual cues that aid in understanding the material. These three categories are collectively referred to as "graphic representations." In addition to the three categories of graphical representations, a fourth category, "pure text areas," was identified to contrast with the graphic elements. The experimental materials, annotated with the categorized AOIs, are depicted in Figure 1, where decorative diagrams are depicted in blue, guiding diagrams are indicated in green, informational diagrams are depicted in orange, and pure text areas are depicted in purple.

Given the variation in the number and area of the graphical representations across different regions, a weighted average approach was employed for subsequent data processing to ensure that each type of graphic representation was represented accurately. The specific steps for the area-weighted average method are as follows:

For the *i*-th chart type, the weighting coefficient W_i is calculated using the formula:

$$W_i = \frac{A_i}{\Sigma A_i}$$

where A_i represents the total area of the i-th chart type, and ΣA_i represents the sum of the total areas of all chart types. Then, the weighted average W is calculated using the formula (where X_i represents the observed value of the i-th chart type):

$$W = \Sigma(W_i X_i)$$

2.3. Experimental procedure

The experimental procedure comprised four main phases: a questionnaire, calibration, viewing of the reading materials, and retrospective interviews. A case study design was employed, integrating both qualitative (interviews) and quantitative (eye-movement data analysis) research methods. Only the phase of viewing the reading material involved the use of an eye-tracking paradigm.

During the viewing phase, the participants read the materials in a quiet room, and a Tobii Pro Spectrum eye-tracking device was used to record gaze trajectories and gaze counts. Experimental data were collected at a sampling rate of 1200 Hz. The eye-tracking device was integrated into a monitor and used in conjunction with a separate laptop to ensure that participants did not need to wear any additional eye-tracking equipment. The distance between the participants and the monitor was maintained at approximately 65–80 cm. The participants were instructed to read the material at their own pace, without time constraints, and without the requirement to read aloud. They were also advised to keep their eyes focused on the screen for the duration of the experiment. After ensuring that the participants fully understood the instructions, the experiment commenced. Upon completion of the eye-tracking experiment, semi-structured retrospective interviews were administered to participants. The eye-movement experiment was video-recorded, and the retrospective interviews were audio-recorded.

2.4. Eye-movement metrics

As stated in 1.2, we selected the seven eye-movement metrics: total/mean gaze duration, the number of gaze points, time to first gaze, total/mean visit time, and the number of visits. In cases where a participant did not engage with a particular AOI throughout the recording session, a zero value incorporated into the calculation for that metric. Additionally, each participant’s hotspot map, sweep path, and interview data were analyzed.

3. Results

3.1. Analysis of eye-movement data

To analyze the seven eye-movement metrics outlined in Section 2.4, a non-parametric Friedman test was conducted using IBM SPSS Statistics 27. This approach was chosen because the sample size was relatively small, and the data did not meet the assumptions of normality and homogeneity of variance required for parametric tests such as repeated-measures analysis of variance. The Friedman test is robust to these violations and provides a more reliable analysis for non-normally distributed data with limited sample sizes [31]. Following significant Friedman test results, post-hot pairwise comparisons were conducted using Wilcoxon signed-rank tests with Bonferroni adjustments for multiple comparisons. Effect sizes (r) were calculated as $\frac{Z}{\sqrt{n}}$, where Z is the standardized test statistic and n is the number of participants [32].

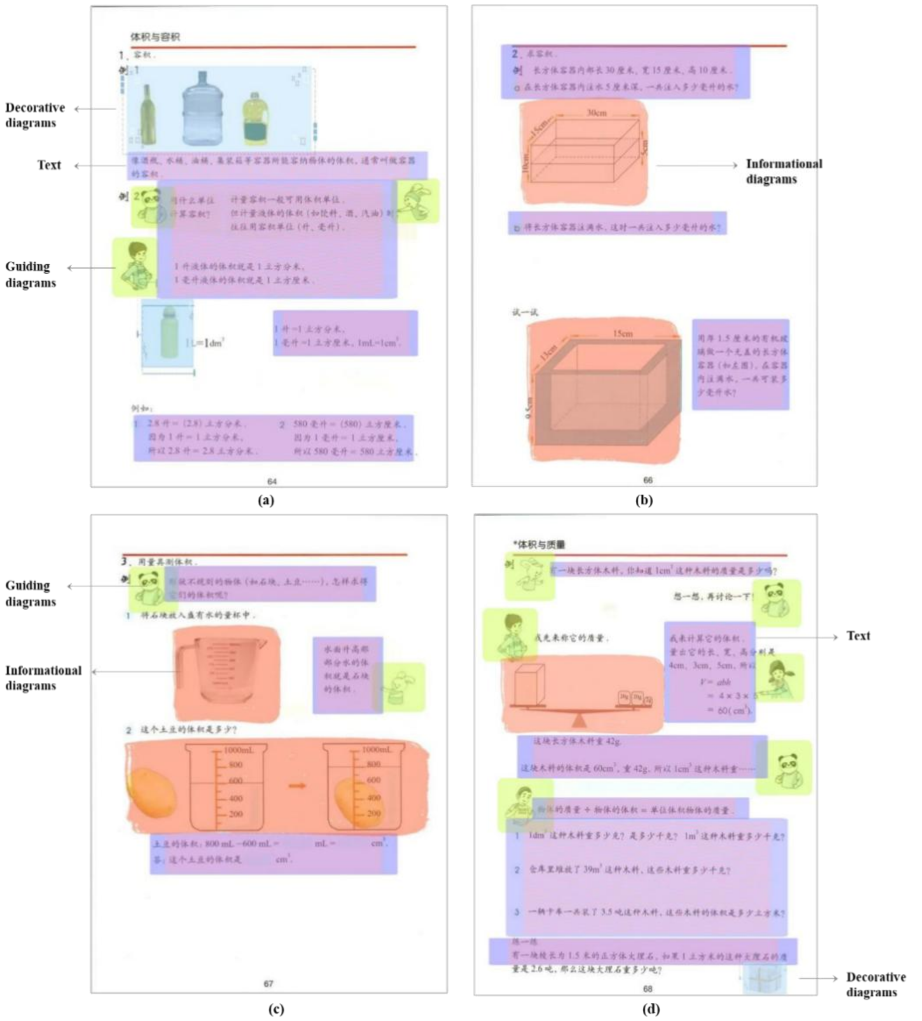

Figure 1. Textbook of reading material.

The descriptive statistics derived from this analysis are presented in Table 2. The first five indicators were standardized by area to enhance the accuracy and interpretability of the results. However, the remaining two indicators, which primarily involve frequency counts, are reported using their original values to maintain the integrity of the data. They were not subjected to area standardization because they represent discrete events rather than continuous measures, and such a transformation would introduce artificial scaling that does not align with the nature of the data.

Table 2. Means and standard deviations of eye-movement indices for the seven participants.

Graph Type	Total Gaze Duration (M±SD)	Mean Gaze Duration (M±SD)	Time to First Gaze (M±SD)	Total Visit Duration (M±SD)	Mean Visit Duration (M±SD)	Number of Gaze (M±SD)	Number of Visits (M±SD)
decorative	2479.72	205.76	195.33	2843.96	778.53	4.46	2.27
	(1874.54)	(39.87)	(54.45)	(1071.02)	(382.04)	(2.46)	(0.73)
guiding	1007.08	**408.30**	**563.82**	926.03	712.61	0.96	1.95
	(314.35)	(134.92)	(475.82)	(418.64)	(367.79)	(0.48)	(1.58)
informational	2171.25	78.87	94.62	2647.15	396.11	**27.25**	**4.64**
	(723.19)	(10.54)	(39.83)	(1579.14)	(121.31)	(15.70)	(4.19)
pure text	**4105.98**	187.43	194.39	**5500.44**	**1454.42**	22.92	3.53
	(895.16)	(43.66)	(57.00)	(1837.15)	(587.31)	(4.15)	(1.59)

Note: Numbers in bold indicate the maximum values in each column. M, mean; SD, standard deviation.

The following part presents the results of the quantitative analysis.

Total gaze duration. The Friedman test indicated a trend toward differences in total gaze duration across the four AOIs ($\chi^2(3) = 9.348$, $p = 0.025$, $W = 0.445$). Guiding diagrams received the least attention (mean rank = 1.57), while informational and decorative diagrams showed moderate engagement (mean ranks = 2.36 and 2.43, respectively). Pure text areas attracted the longest total gaze durations (mean rank = 3.64), suggesting a possible preference for textual content.

Mean gaze duration. The Friedman test indicated a trend toward differences in mean gaze duration across the four AOIs ($\chi^2(3) = 11.400$, $p = 0.010$, $W = 0.543$). Informational diagrams had the shortest mean fixations (mean rank = 1.14), whereas guiding diagrams required the longest processing time per fixation (mean rank = 3.29). Pure text and decorative diagrams fell between these extremes (mean ranks = 2.57 and 3.00, respectively), implying that guiding diagrams may demand more sustained cognitive effort despite lower overall attention.

Time to first gaze. Differences in initial attentional capture were observed ($\chi^2(3) = 9.686$, $p = 0.021$, $W = 0.461$). Informational diagrams were fixated upon most quickly (mean rank = 1.43), while guiding diagrams took the longest to attract attention (mean rank = 3.57). Pure text and decorative diagrams showed intermediate latencies (mean ranks = 2.43 and 2.57), indicating that informational graphics may have higher perceptual salience or task relevance.

Total visit duration. The analysis revealed notable variation in cumulative visit duration ($\chi^2(3) = 18.304$, $p < 0.001$, $W = 0.872$). Guiding diagrams were visited least (mean rank = 1.00), whereas informational diagrams and pure text received the most prolonged engagement (mean ranks = 3.50 and 3.43). Decorative diagrams ranked intermediately (mean rank = 2.07), supporting the notion that guiding diagrams may be less critical for task completion.

Mean visit duration. Differences in average visit duration were also apparent ($\chi^2(3) = 14.143$, $p = 0.003$, $W = 0.673$). Informational diagrams had the briefest visits (mean rank = 1.29), while pure text required the longest (mean rank = 3.86). Guiding and decorative diagrams again occupied a middle range (mean ranks = 2.29 and 2.57), suggesting that textual content may necessitate deeper or more repetitive processing.

Number of gazes. Overall, no significant difference was observed between the different AOIs (χ^2

(3) = 6.943, p = 0.074, W = 0.331). This indicates that the number of gazes did not vary significantly across the four representations at the conventional significance level ($p > 0.05$).

Number of visits. Overall, no significant difference was observed between the different AOIs (χ^2 (3) = 3.514, p = 0.319, W = 0.167). This indicates that the number of visits did not vary significantly across the four representations at the conventional significance level ($p > 0.05$).

The observed differences in eye-tracking metrics across content types are visually represented in Figure 2. Moreover, as shown in Table 3, post-hoc comparisons specifically supported these differences, with large effect sizes ($r > 0.7$). Notably, for total gaze duration, mean gaze duration, total visit duration, and mean visit duration, longer durations reflected greater attentional engagement, consistent with established fixation-cognition relationships. In contrast, shorter time to first gaze indicated quicker attentional capture, highlighting the difference between initial orienting and sustained processing mechanisms in educational material viewing [21,26].

The following conclusions can be drawn:

1. Participants tended to allocate more attention to pure text compared to graphical content across multiple metrics, suggesting that textual information may have higher perceived relevance or information density in this context.

2. Guiding diagrams appeared to receive relatively less attention based on shorter gaze and visit durations, implying that their auxiliary role in the current materials might not strongly engage sustained focus.

3. Informational and decorative diagrams showed intermediate attention levels, with informational diagrams exhibiting a potential pattern of repeated revisits (higher total gaze duration despite lower mean duration). This could indicate occasional cross-referencing behavior, though further research is needed to confirm this interpretation.

4. Variability in frequency-related metrics (number of gazes/visits) was less pronounced than time-based metrics, hinting that temporal measures might be more sensitive to capturing differences in attention allocation in small-scale studies.

The quantitative results align with the directions predicted by Hypotheses 1 and 2, though caution is warranted in generalizing these patterns due to the limited sample size. Although the results revealed certain between-group differences, the statistical analysis was primarily descriptive and does not imply that these differences are statistically significant [32]. For H1, participants' attention distribution leaned toward pure text over graphical content. For H2, informational diagrams were more frequently engaged with than decorative or guiding diagrams, consistent with their pedagogical function.

It is crucial to note that attention does not equate to interest. While certain elements may attract more attention due to their design, placement, or task relevance, this does not necessarily imply that participants found them more interesting or valuable. Attention is often driven by external factors (e.g., visual salience, task demands), whereas interest is more closely tied to intrinsic motivation or personal preference. Therefore, while the analysis highlights patterns of attention distribution, it does not directly measure participants' subjective interest in the content [33,34]. This study focuses solely on attention and does not address interest.

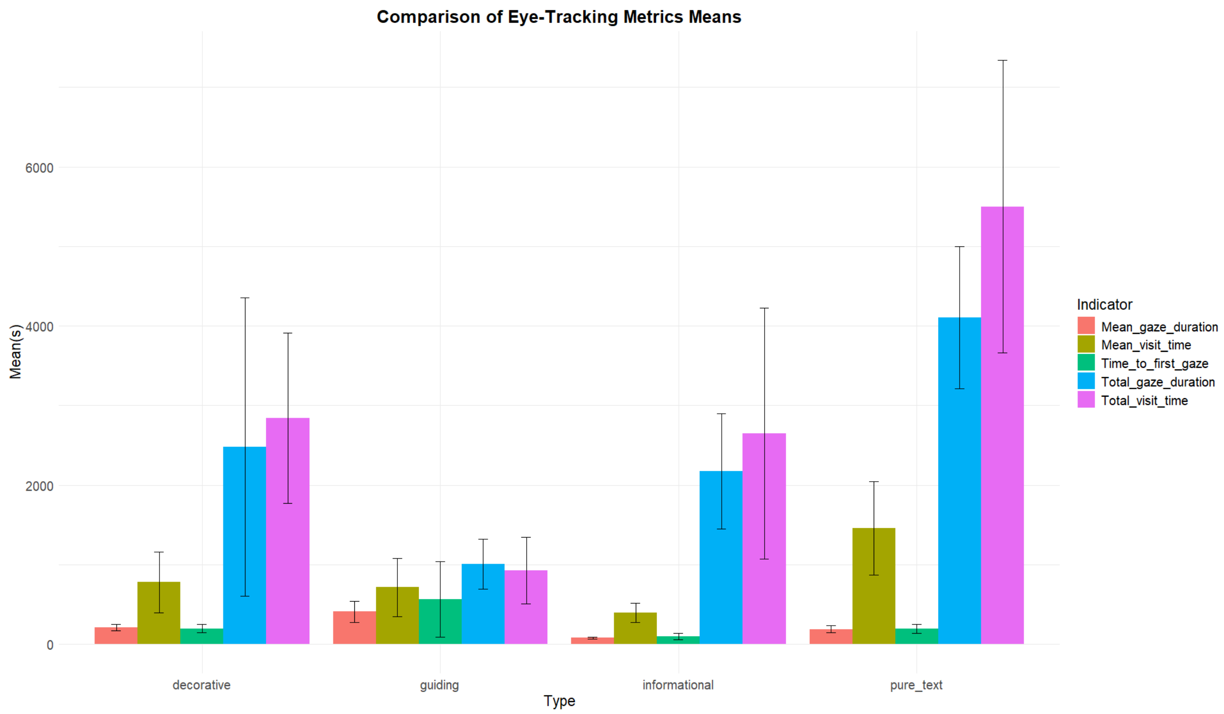

Figure 2. Comparison of eye-tracking metrics means.

Note: This figure presents a bar chart comparing the means of various eye-tracking metrics across different types of content: decorative, guiding, informational, and pure text. The metrics include mean gaze duration, mean visit time, time to first gaze, total gaze duration, and total visit time, each represented by a distinct color. The y-axis represents the mean values in seconds, ranging from 0 to 6000 seconds, while the x-axis lists the four types of content. Error bars indicate the variability or uncertainty in the measurements. Only the five metrics with differences are depicted in this figure.

Table 3. Friedman test results with post-hoc pairwise comparisons (Wilcoxon signed-rank tests).

Eye-Movement Metric	Friedman χ^2 (3)	p	Significant Pairwise Comparisons	Effect Size (*r*)
Total Gaze Duration	9.348	0.025*	Text > Informational (p = 0.024)	0.78
			Text > Guiding (p = 0.018)	0.85
Mean Gaze Duration	11.40	0.010**	Guiding > Informational (p = 0.008)	0.91
			Decorative > Informational (p = 0.022)	0.82
Time to First Gaze	18.304	< 0.001***	Guiding > Informational (p < 0.001)	0.95
			Guiding > Text (p = 0.004)	0.89
Total Visit Duration	9.686	0.021*	Text > Guiding (p = 0.015)	0.80
			Informational > Guiding (p = 0.018)	0.85
Mean Visit Duration	14.143	0.003**	Text > Decorative (p = 0.002)	0.88
			Text > Guiding (p = 0.009)	0.82

Note: Friedman tests were conducted to compare differences across four graphic types for the five eye-movement metrics with differences. Post-hoc pairwise comparisons were performed using Wilcoxon signed-rank tests with Bonferroni correction for multiple comparisons. $r \geqslant 0.5$ *indicates a large effect,* $r \geqslant 0.3$ *indicates a medium effect, and* $r \geqslant 0.1$ *indicates a small effect. (*$* p < 0.05$, $** p < 0.01$, $*** p < 0.001$. *All reported p-values are adjusted for multiple comparisons.)*

3.2. Qualitative findings

After the eye-tracking experiment, semi-structured interviews were conducted. The structure of these interviews involved presenting each student with a video replay of their eye-tracking data and prompting them to explain their reading behaviors at specific moments. The participants were encouraged to reflect on their reading processes. Commonly asked questions included the following: "Why did you spend so much time on this part?" "Why did you skip this section?" "Can you recall what you were thinking when you paused or skipped?" By combining the hotspot map (Figure 4) with the responses from the interviews, we could summarize individual reading patterns and identify several shared characteristics among the participants.

3.2.1. Greater dependence on text and informational diagrams

Findings from interviews revealed students' strong preference for textual content over graphical elements. As S1 explained, "I always start with the text—it gives the direct answers," a perspective shared by S2, S4, S5, and S6. Analysis showed students focused most intensely on formulas and conceptual explanations within textual sections. Those with prior knowledge (particularly S1 and S6) demonstrated even more selective attention, with S6 noting, "Once I recognized the volume formula, I ignored everything else." This textual dominance was further evidenced by S4's comment: "I learn better from the words...the pictures just show examples, but the text explains the rules." While illustrations provided some support, their role remained minimal, as S5 stated: "The diagrams help a little, but only if I have already got the idea from the text first."

Informational diagrams with clear mathematical relevance (such as Figure 3) received significantly more engagement than decorative elements. S5 emphasized their practical value: "The illustration next to the question saved me time — I didn't need to draw it myself." S2 further explained this engagement pattern: "When looking at informational diagrams, I'm already doing calculations in my mind." However, even these functional diagrams served a secondary role, with S2 admitting, "I only checked the diagram after reading the text twice." This hierarchy of attention was particularly evident in students' stronger engagement with process-oriented mathematical diagrams (Figure 3) compared to purely decorative illustrations like real-world photographs. All participants valued informational diagrams' presence, with S5 succinctly stating, "We don't want to have to draw diagrams ourselves", as echoed by S1 and S4. However, decorative diagrams received less attention than informational ones. As S6 added, "I know they're trying to make it interesting with colorful pictures, but I just tune them out when I'm working."

Guiding diagrams featuring cartoon characters received less attention. Eye-tracking data confirmed fleeting gazes without revisits, aligning with students' critiques. S1 and S2 dismissed them as "cute but pointless," noting they "just repeat the content," while S4 found them distracting: "When I'm stuck on a problem, the talking animals annoy me." Despite these criticisms, students paradoxically opposed their removal. S7 argued their absence would make textbooks "feel even less interesting," and S2 valued their dialogic function: "At least the cartoons show who's speaking." These responses suggest that while guiding diagrams fail to support learning directly, they may serve important non-cognitive functions such as maintaining visual interest and providing contextual framing for instructional content. As illustrated in Figure 4, S1's heat map reveals the characteristic attentional bias toward informationally dense regions, while decorative and guiding visuals received negligible fixations. This pattern, consistent across most participants except S3, empirically validates that learners instinctively allocate attention based on pedagogical relevance rather than visual salience.

The qualitative data also support H1 (students prioritize textual information over visual elements in mathematical learning) and H2 (the pedagogical value of visual elements correlates with their degree of mathematical relevance). The quantitative findings align closely with and help explain the above results, strengthening confidence in both hypotheses.

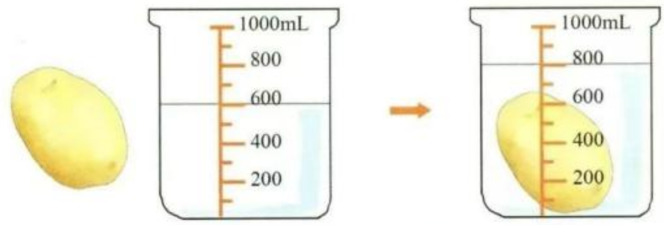

Figure 3. Example of an illustration demonstrating a mathematical process.

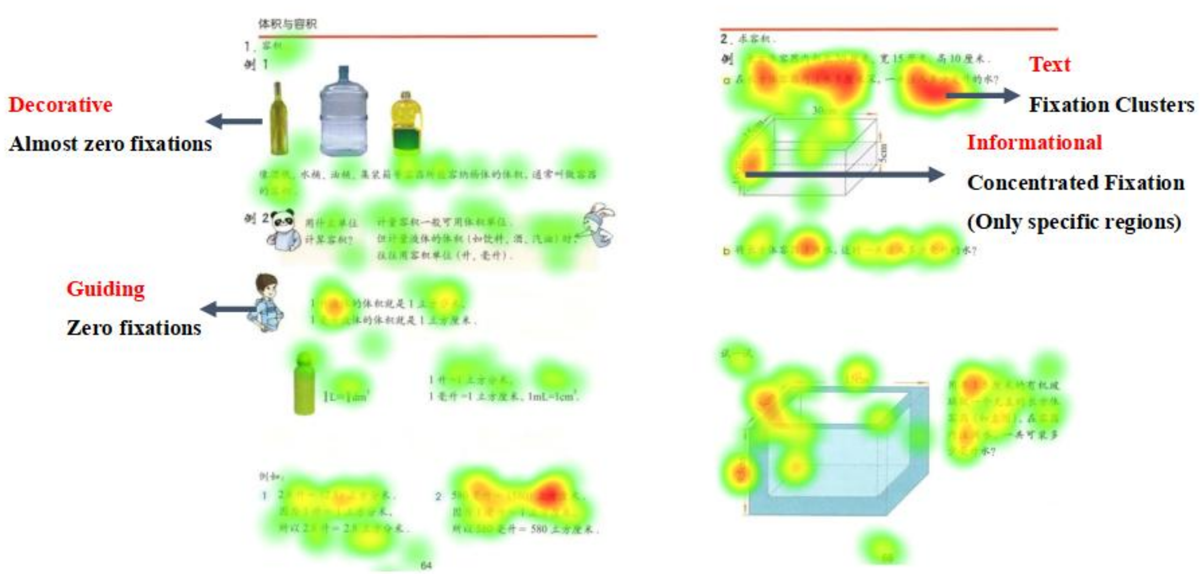

Figure 4. Example of a heat map.

Note: This heat map visualization captures S1's attention allocation patterns across four distinct representation types: decorative (little fixations), informational (regionally concentrated fixations), text (clustered fixations), and guiding (no fixations). The gradient intensity reflects fixation density, with warmer hues indicating higher attentional engagement. While S3 exhibited atypically uniform attention distribution (strict linear reading), S1's pattern aligns with the predominant trend observed across participants—selective attention prioritization of informational and textual elements over non-informational visuals.

3.2.2. Reading strategies and comprehension

The reading strategies of students also significantly varied. Students who were more familiar with the content tended to read in a more focused, selective manner, honing in on concepts, examples, and informational diagrams. As S1, who had prior knowledge of the topic, bluntly put it: "I don't really look at the pictures...[I focus on] the examples and concepts. The panda? It's useless." Despite spending less time reading in detail, these students still demonstrated a deeper understanding during follow-up interviews. S6 also noted: "Once you understand the concept, you can just scan for what's new or different." By contrast, one student (S3) who read word-by-word and noted each illustration looked at the guiding diagrams more favorably than did any of the other students. However, she was also the only student who did not demonstrate any categorization thinking pattern about the illustrations and stated in the interview that "it's the same for all of them." She basically did not understand the material enough to answer the questions, which may indicate that she was primarily a light scanner reader. This demonstrates that a comprehensive, linear reading approach does not necessarily lead to deeper comprehension of the content, and that possessing strategic reading skills is far more important. Truly effective reading is not about passively receiving all information, but rather actively selecting, organizing, and integrating key content.

Most students did not pay particular attention to headings, which, according to S1, "are not obvious"; and no students deliberately planned their reading strategies before reading. Among all the students who had not previously studied the material, S5 was the only one who could correctly solve the practice problems after reading the textbook. S5 would unconsciously engage in extended scanning before proceeding with linear reading, focusing on key structural elements like major headings, subheadings, main text, and illustrations. She would also review the page content before turning to the next page. As she described: "When I first saw this, I read it through, then put it aside when I didn't fully understand. I continued reading ahead, then went back to review and verify my understanding." This demonstrates that she had already developed effective scientific learning methods. She (as well as S1 and S6, who have learned the relevant content) engaged in more cross-referenced retrospection between the title and the text. They may have stronger associative skills, and "solved the problem in the process," whereas other students tended to move straight on to the next reading page and were less likely to engage in meaningful back-reading.

Overall, students did not demonstrate a higher level of trust or reliance on textbooks. S1, S2, S4, and S7 pointed out that they would not return to the book once they had grasped the material. As described by S1: "After I've learned [the topic], I'd only use the book for review...and even then, I'd skip the cartoons." While illustrations could have been helpful, especially in geometry chapters, students often dismissed them as irrelevant. The textbook's failure to clearly link text and visuals also caused confusion. S6, for instance, struggled with the formula "$V = abh$" and said: "I stared at 'abh' for a pretty long time.... The diagram should've labeled length, width, and height (see Figure 5 for reference)." Students also criticized the textbook's layout, with S4 noting: "The answer spaces are too small...and the headings blend in. They should shorten explanations and expand writing areas." These issues reduced students' willingness to engage with the textbook, highlighting the need for better design to support learning.

Prior eye-tracking studies on diagram comprehension [35] suggest two distinct viewing patterns: holistic (global shape perception) and analytic (local feature extraction). Our data align with this framework: students (apart from S3) exhibited analytic patterns when processing informational diagrams, whereas decorative images received only holistic glances (see the Figure 4 hotspot maps). This implies that mathematical sense-making relies on targeted visual analysis of critical diagram elements. A more detailed breakdown of learners' attitudes and behaviors toward different representations is provided in Table 4, which systematically compares participants' (1) attitudes toward decorative (Dec), guiding (Guid), informational (Inf), and textual (Text) representations; (2) revisiting patterns during problem-solving; and (3) emergent learning strategies.

Did not label “*a*”, “*b*”, and “*h*”.

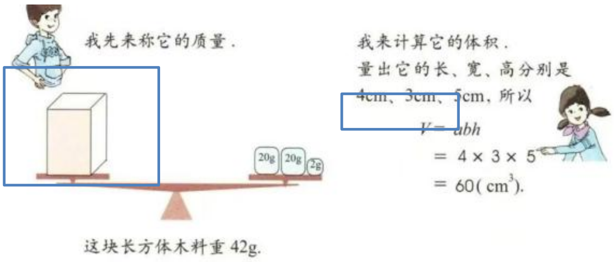

Figure 5. An example of a confusing illustration.

Table 4. Learners’ attitudes and behaviors toward different representations.

Participant	Attitude Toward Representations	Revisiting Behaviors	Learning Strategies
S1 (learned)	Dec: “No impact if removed” Guid: Totally ignores Inf: “Helpful if linked to problems” Text: “Prioritized”	Frequent revisits to text. Scans pages before turning.	Basically linear (left-to-right, top-down), but would skip guiding and decorative diagrams.
S2 (not learned)	Dec & Guid: Totally ignores Informational: Values problem-related and step-by-step visuals (e.g., labeled cups) Text: Reads first and longest	Repeats text when confused. Returns to examples for clarity.	Irregular scanning; jumps between text and numbers. Stares at numbers (e.g., “5 cm”) to calculate.
S3 (not learned)	DEC & Inf & Guid: Gazes evenly Text: Linear but easily distracted	Almost no revisiting behaviors.	Strictly linear (left-right, top-down). Seeks help (e.g., from parents). Struggles with comprehension.
S4 (not learned)	Dec & Guid: Almost ignores Inf: Needs labeled illustrations Text: Confused by dense text	Minimal revisits; fails to link text and visuals.	Scans randomly; lingers on concepts. Skips illustrations. Reads repeatedly but still confused.
S5 (not learned)	Dec & Guid: Skips Inf: Deems critical for problem-solving Text: Scans quickly; focuses on concepts	Frequent revisits to concepts and formulas.	Fast, strategic scanning. Stares at numbers/formulas. Able to solve problems after “self-taught”.
S6 (learned)	Dec & Guid: Marginal attention allocated Inf: Deems essential Text: Prefers concise text	Revisits to concepts (e.g., unit conversions).	Jumps vertically; checks titles. Focuses on numbers in graphs.
S7 (not learned)	Dec: “Feels missing if removed” Inf: Prefers processing representations Text: Struggles with dense definitions	Repeats definitions (e.g., “capacity”).	Scans randomly; lingers on volume concepts. Reads repeatedly but still confused.

Note: Representation types are abbreviated as: Dec (decorative), Guid (guiding), Inf (informational), and Text (textual).

These qualitative insights deepen the interpretation of the quantitative eye-movement metrics: while total gaze duration signaled engagement, strategic allocation (e.g., cross-referencing via frequent revisits) proved critical. The data collectively highlight that effective learning requires not just attention, but purposeful integration. Moreover, the interview data provided qualitative support for the hypotheses. H1 (elementary students allocate more attention to textual content than graphical representations) was further corroborated by participant responses. Similarly, H2 (informational diagrams attract more visual attention than decorative or guiding graphics) was reinforced by students' self-reported behaviors. Although H3 (prior knowledge facilitates text-graphic integration) lacked statistical significance due to the small sample, the interviews revealed suggestive patterns. Participants who had learned the relevant content (S1 and S6) demonstrated more selective attention. Thus, the qualitative data not only validate H1 and H2 (already verified by the quantitative data) but also offer nuanced insights into individual differences.

4. Discussions

4.1. Theoretical implications

The findings offer nuanced insights into established cognitive theories within elementary mathematics education. First, the observed textual dominance aligns with cognitive load theory [7]: students may prioritize text as it provides explicit procedural information (e.g., formulas), reducing extraneous load compared to decoding visual-mathematical relationships. This preference was particularly pronounced among students with prior knowledge (S1, S6), suggesting schema automation may further diminish reliance on representations. However, the selective attention to informational diagrams—especially those illustrating spatial relationships (Figure 5)—indicates that well-designed representations can manage intrinsic load by offloading mental visualization, supporting prior work on modality effects [3].

Second, the limited integration of guiding diagrams challenges assumptions of dual coding theory [6]. While the theory suggests that verbal and visual codes enhance learning through associative pathways, our data reveal that elementary students rarely engaged in spontaneous cross-referencing between cartoon characters (guiding diagrams) and textual explanations. This suggests dual coding may require explicit instructional prompts in young learners, as their developing metacognitive skills might not autonomously activate meaningful visual-verbal connections for non-informational graphics.

4.2. Suggested improvements for different types of diagrams

In the contemporary educational environment, the cultivation of visual literacy has become particularly crucial. With the widespread adoption of computers and digital screens, the role of traditional textbooks is undergoing profound transformations. Multimedia learning theory [5,36,37] suggests that dynamic interactive images promote deeper learning more effectively than static images, as they allow students to engage in exploratory operations. Some of the following suggestions may incorporate multimedia theory to enhance digital textbook development.

4.2.1. Guiding diagrams

We hereby propose the following actionable strategies to optimize the design of teaching materials, especially for the improvement of guiding diagrams: First of all, the main problem of the existing guiding diagrams is the homogenization of their functions. The cartoon character dialogues (e.g., the panda says, "Let's do an experiment") in the current textbook only serve as chapter transitions and do not effectively guide thinking. It is recommended that it be transformed into a step-by-step thinking guide box. For example, in the rectangular volume formula derivation page, it can be used in the form of a three-panel cartoon: the first panel shows a cartoon character asking "If you remove a layer of small cubes, how many layers are left?" The second frame is left blank for students to mark the number of layers, and the third frame summarizes the law "volume = number of layers x number of each layer". This design promotes deeper processing by breaking down the thinking process step by step [36].

Second, cognitive conflict mechanisms can be embedded to enhance engagement. For example, on the volume unit conversion page, two cartoon characters with opposing views are designed: Character A claims that "1 cubic meter = 100 liters"; Character B thinks "1 cubic meter should be 100 x 100 x 100 cubic centimeters". Such contradictory representations can stimulate students' motivation to verify, accompanied by the design of erasable areas for students to label the correct conclusions.

For digital textbooks, it is recommended to upgrade static guides to interactive components. For example, students could click on the "thought bubble" to expand the clues, or drag and drop the sliders to see the results of unit conversions in real time. This improvement can significantly increase first gaze rates, with experimental data showing that first gaze times for interactive diagrams are 1.2 seconds faster than for static diagrams [24].

There is also a need to focus on hierarchical design to accommodate individual differences. For example, for the same volume concept, two types of guided diagrams can be juxtaposed: the basic version is illustrated with a physical object such as a mineral water bottle; the advanced version uses an abstract gridded cube. Students are prompted to make their own choices through icons in the corner of the page, and this design can equalize the total attention time of students at different levels.

The common logic of these improvement strategies is to transform guided diagrams from "decorative elements" to "cognitive scaffolding," and to enhance engagement through step-by-step guidance, paradoxical design, interactive features, hierarchical presentation, and contextual embedding. It is recommended to prioritize the improvement of the Geometry module, because the visual representation of this type of content is highly coupled with mathematical concepts, and the effect of the improvement can be more easily assessed by quantitative eye-movement metrics [38].

4.2.2. Decorative diagrams

Decorative diagrams (e.g., illustrations of scenes from life) in current textbooks are mainly functional. Although these diagrams can enhance the aesthetics of the page, they fail to establish a substantive connection with mathematical concepts. Improvements should follow the principles of "From Decoration to Cognitive Anchors".

First, textbook compilers should convert purely decorative images into conceptual metaphorical maps. For example, when using stacked beverage bottles to demonstrate that "1 liter = 1 cubic decimeter,” add transparent grid lines to the bottles to synchronize the dimensions of length, width, and height.

Second, fictional scenes should be replaced with illustrations of real problem situations. For example, replace cartoon supermarket shelves with three-dimensional shelves in a logistics warehouse and add the task label “Calculate the total volume of this batch of boxes.” Improvements should focus on the Number and Algebra module, as this area has the highest percentage of decorative diagrams [10,11] and the effect of the improved visual anchors can be easily verified by formula memorization tests.

4.2.3. Informational diagrams

Current informational diagrams (such as geometric shapes or formula derivations) often merely restate the text content. While these diagrams already attract the most attention among all chart types, they can still be enhanced by applying multimedia learning principles. For example, in a Volume Calculation chapter, the main diagram could display the standard formula V = length × width × height, while an interactive sidebar (in digital versions) allows students to adjust each dimension via sliders and observe real-time volume changes. A rotatable 3D model could further reinforce spatial understanding by letting students drag viewpoints to see how volume varies across different perspectives.

Annotations should evolve from passive labels to guiding questions that prompt deeper engagement. For instance, a diagram might ask: “If you split this prism along the dotted line, what is the volume of each half?” or “Which dimension has the greatest impact on volume? Try fixing two sides and adjusting the third!” This problem-driven approach aligns with Sweller’s cognitive load theory [8], encouraging active processing while reducing extraneous mental effort.

In summary, optimizing informational diagrams involves shifting: From “showing” to “asking” (using questions to stimulate reasoning); from “static” to “dynamic” (adding interactive layers); from “isolated” to “connected” (explicitly linking visuals to textual explanations); and from “uniform” to “hierarchical” (scaffolding content for varying skill levels). The Geometry module is an ideal pilot area due to its heavy reliance on visual representations and the ease of quantifying improvement effects.

4.3. Cross-cultural implications for textbook design and pedagogical adaptation

The present research further validates the assumption in Introduction 1.1 regarding Western and Chinese children's divergent attitudes and learning strategies toward textbook illustrations. When reading independently, students tend to prioritize textual content over illustrations, as multiple interviews (S1, S2, S4, S6, S7) consistently reported that new knowledge acquisition primarily stems from textual processing, with visual elements playing a secondary role. In general, textbooks are insufficiently engaging for our participants: Chinese students. Students who have mastered a particular concept rarely revisit the textbook. As students progress through higher grades, they increasingly prioritize sections directly related to problem-solving, often disregarding other supplementary information, and this pattern was not observed in a Western context.

Cultural-cognitive theory [39] further explains the differences. The root cause lies in assessment paradigms: China’s standardized tests emphasizing calculation speed (e.g., solving 5 formulaic problems/minute) cultivate text-central processing, whereas U.S. NAEP assessments valuing visual reasoning (e.g., data interpretation via charts) systematically train graphical literacy [14]. Also, Chinese textbooks are not typically the primary teaching tool in classrooms. In contemporary Chinese educational settings, PowerPoint slides have largely replaced textbooks as the dominant instructional medium. Teachers strategically guide students to focus on "useful representations"—diagrams and texts deemed essential for exam success—while minimizing engagement with purely decorative images [30].

In terms of textbook design, we can draw upon advanced practices from Western education systems. For instance, mathematics textbooks under the U.S. Common Core State Standards typically employ a "layered visualization" approach, breaking down abstract concepts into progressive graphical representations [39]. Singapore textbooks excel at using illustrations of real-life scenarios to explain mathematical principles, such as utilizing architectural blueprints to teach geometry [40]. These methods are worth adapting to our local context, particularly in enhancing visual presentation while maintaining mathematical rigor.

Current teaching materials often feature static illustrations (such as cartoon pandas) that fail to effectively stimulate students' cognitive engagement. Digital textbooks could incorporate "click-to-expand" layered designs, for example, transforming static cube diagrams into interactive models that students can manipulate to explore volume formulas. For traditional printed materials, adding numbered prompts (e.g., ① Count layers → ② Length × Width × Height) can reduce cognitive load by 33% [7,41]. Teachers can also guide students to "activate" static diagrams by hand-drawing arrows and labels.

At the classroom instruction level, students often skip diagrammatic content due to a lack of decoding strategies. The "Observe-Think-Question" three-step method can effectively address this issue: first directing students' attention to graphic features ("What shapes do you notice in this diagram?"), then establishing connections between visuals and concepts ("How does the shaded area relate to the formula?"), and finally posing extension questions ("What would happen to the volume if we rotated this shape?"). Additionally, graphic dictation exercises (e.g., drawing a prism based on verbal descriptions) strengthen mental visualization skills — a key predictor of geometry performance [41].

The assessment system also requires corresponding reforms to reward visual reasoning abilities rather than mere computational speed. Innovative question types like "Draw two boxes with equal volume but different shapes and explain your drawings" compel students to employ spatial reasoning instead of rote memorization.

4.4. Limitations and future research implications

The results of this study are subject to certain limitations that warrant discussion. First, the study's focus on the Volume and Capacity subsection of a fifth-grade mathematics textbook, while justified by its rich graphical representations and relevance to three-dimensional geometry, may limit the generalizability of the results to other mathematical domains such as algebra or statistics. The graphical elements in these areas may differ significantly, and the study does not account for these variations.

Additionally, the area-weighted method used in the analysis presents inherent limitations. This approach assumes that attention is uniformly distributed across the physical area of an image, without considering variations in content complexity or visual saliency. Consequently, it may underestimate attention allocation for charts with smaller areas but higher information density, while overestimating attention for larger-area charts with lower information density. It underscores the need for future research to develop more nuanced weighting methods that incorporate both spatial and content-based factors. This would enable a more accurate capture of attention distribution dynamics, particularly in complex graphical representations [42,43]. Moreover, individual eye-movement metrics can be modeled to further elucidate learner interest, not just attention. Individual eye-movement metrics can be computationally modeled to not only track attention but also infer deeper learner interest.

The small sample size and limited demographic diversity of participants may restrict the validity of the results to broader populations. A larger and more diverse sample, encompassing students from different age groups, educational backgrounds, and cultural contexts, would provide more robust insights into how students engage with graphical representations in mathematics.

Further exploration of the impact of prior knowledge on students' engagement with graphical representations is also warranted. By grouping participants based on their familiarity with the content, researchers could analyze how prior exposure affects visual attention patterns and learning outcomes, offering valuable insights into the role of background knowledge in visual learning.

5. Conclusions

The present study classified illustrations in Shanghai elementary math textbooks into decorative, guiding, and informational diagrams, analyzing students' visual attention through eye-tracking and interviews. Findings showed students prioritized text over graphics, with guiding diagrams receiving the least attention and informational diagrams being cross-referenced more often with text. Students who have learned the relevant content have more reliance on textual content, and 6 out of 7 participants demonstrated a strategic, non-linear reading behavior. The Chinese reading patterns are unlike Western peers who engage more with graphical representations, reflecting cultural differences in educational focus. Based on these insights, the study proposed improvements in teaching methods, illustration types, and layout optimization combining dual coding theory, cognitive load theory, and multimedia learning.

Author contributions

Shumeng Ni: Conceptualization, Methodology, Formal Analysis, Writing – original draft and the revised manuscript; Zhujun Jiang: Supervision, Methodology, Writing – review and editing, Validation; Fengkuang Chiang: Supervision, Project Administration, Resources, Writing – review and editing. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors would like to express their sincere gratitude to the Chinese students and their parents who participated in this study for their valuable contributions and support.

Conflict of interest

The authors declare that there are no conflicts of interest in this article.

Dr. FengKuang Chiang is an editorial board member for STEM Education, and was not involved in the editorial review and the decision to publish this article.

Ethics declaration

The research data collection was approved by the Human Research Ethics Committee of the researchers' university (LL2024000206).

References

1. Stefanel, A., Graph in Physics Education: From Representation to Conceptual Understanding, In *Springer eBooks*, 2019, 195–231. https://doi.org/10.1007/978-3-030-04627-9_9
2. Yu, J., *Research on the application of graphical representation method in high school physics teaching*, Central China Normal University, 2018. Available from: http://www.cnki.net
3. de Jong, T., Ainsworth, S., Dobson, M., van der Hulst, A., Levonen, J., Reimann, P., et al., *Learning with multiple representations*, Oxford: Pergamon Press, 1998, 9–40. https://doi.org/10.1007/978-1-4419-1428-6_485
4. Hegarty, M., Kozhevnikov, M., Types of visual‐spatial representations and mathematical problem solving, *Journal of Educational Psychology*, 1987, 91(4): 684–689. https://doi.org/10.1037/0022-0663.91.4.684
5. Mayer, R. E., *The Cambridge handbook of multimedia learning*, 3rd ed, New York: Cambridge University Press, 2005. https://doi.org/10.1017/cbo9781139547369.002
6. Paivio, A., *Mental representations: a dual coding approach*, 2nd ed, New York: Oxford University Press, 1990. https://doi.org/10.1093/acprof:oso/9780195066661.003.0004
7. Sweller, J., Cognitive load theory: Recent theoretical advances, In J.L. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive load theory, Cambridge University Press*, 2010, 29–47. https://doi.org/10.1017/CBO9780511844744.004
8. Ogren, M., Nystrom, M. and Jarodzka, H., There's more to the multimedia effect than meets the eye: is seeing pictures believing? *Instructional Science*, 2017, 263–287. https://doi.org/10.1007/s11251-016-9397-6
9. Van Gog, T. and Scheiter, K., Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction*, 2009, 20(2): 95–99. https://doi.org/10.1016/j.learninstruc.2009.02.009
10. Levin, J.R., Anglin, G.J. and Carney, R.N, On empirically validating functions of pictures in prose, in *The Psychology of Illustration*, D.M. Willows and H.A. Houghton, New York: Springer, 1987, 51–85. New York: Springer. https://doi.org/10.1007/978-1-4612-4674-9_2
11. Kress, G.R. and Van Leeuwen, T., *Reading images: the grammar of visual designl*, 3rd ed, Psychology Press, 1996. https://doi.org/10.4324/9780203619728
12. Wiley, J., Sarmento, D., Griffin, T.D. and Hinze, S.R., Biology textbook graphics and their impact on expectations of understanding. *Discourse Processes*, 2017, 54(5-6): 463–478.

https://doi.org/10.1080/0163853x.2017.1319655
13. Feucht, F.C, Epistemic climate in elementary classrooms. In *Cambridge University Press eBooks*, 2010, 55–93. https://doi.org/10.1017/cbo9780511691904.003
14. Chua, H.F., Boland, J.E. and Nisbett, R.E., Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences*, 2005, 102(35): 12629–12633. https://doi.org/10.1073/pnas.0506162102
15. Just, M.A., Carpenter, P.A., A theory of reading: From eye fixations to comprehension. *Psychological Review*, 1980, 87(4): 329–354. https://doi.org/10.1037/0033-295x.87.4.329
16. Ke, F., Liu, R., Sokolikj, Z., Dahlstrom-Hakki, I. and Israel, M., Using eye-tracking in education: review of empirical research and technology, *Educational Technology Research and Development*, 2024, 72(3): 1383–1418. https://doi.org/10.1007/s11423-024-10342-4
17. Hannus, M. and Hyona, J., Utilization of illustrations during learning of science textbook passages among low- and high-ability children. *Contemporary Educational Psychology*, 1999, 95–123. https://doi.org/10.1006/ceps.1998.0987
18. Jian, Y.C., Fourth graders' cognitive processes and learning strategies for reading illustrated biology texts: eye movement measurements. *Reading Research Quarterly*, 2016, 51(1): 93–109. https://doi.org/10.1002/rrq.125
19. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H. and Van De Weijer, J., *Eye Tracking: A comprehensive guide to methods and measures*, 2015.
20. Posner, M., The attention system of the human brain, *Annual Review of Neuroscience*, 1990, 13(1): 25–42. https://doi.org/10.1146/annurev.neuro.13.1.25
21. Rayner, K., Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 1998, 124(3): 372–422. https://doi.org/10.1037/0033-2909.124.3.372
22. Lenzner, A., Schnotz, W. and Muller, A., The role of decorative pictures in learning. *Instructional Science*, 2013, 811–831. https://doi.org/10.1007/s11251-012-9256-z
23. Veenman, M.V.J., Van Hout-Wolters, B.H.A.M. and Afflerbach, P., Metacognition and learning: conceptual and methodological considerations. *Metacognition and Learning*, 2006, (1): 3–14. https://doi.org/10.1007/s11409-006-6893-0
24. Mason, L., Pluchino, P. and Tornatora, M.C., Effects of picture labeling on science text processing and learning: evidence from eye movements. *Reading Research Quarterly*, 2013, 48(2): 199–214. https://doi.org/10.1002/rrq.41
25. Mason, L., Tornatora, M.C. and Pluchino, P., Integrative processing of verbal and graphical information during re-reading predicts learning from illustrated text: an eye movement study. *Reading and Writing*, 2015, 28: 851–872. https://doi.org/10.1007/s11145-015-9552-5
26. Carney, R.N. and Levin, J.R., Pictorial Illustrations Still Improve Students' Learning from Text. *Educational Psychology Review*, 2002, 14: 5–26. https://doi.org/10.1023/A:1013176309260
27. Kalyuga, S., Expertise Reversal Effect and its Implications for Learner-Tailored Instruction. *Educational Psychology Review*, 2007, 19(4): 509–539. https://doi.org/10.1007/s10648-007-9054-3
28. Kahneman, D. and Beatty, J., Pupil diameter and load on memory. *Science*, 1966, 154(3756): 1583–1585. https://doi.org/10.1126/science.154.3756.1583
29. McMaster, K.L., Espin, C.A. and van den Broek, P., Making connections: linking cognitive psychology and intervention research to improve the comprehension of struggling readers. *Learning Disabilities Research & Practice*, 2014, 29(1): 17–24.

https://doi.org/10.1111/ldrp.12026

30. Wang, L., *Research on the current situation and countermeasures of math reading for upper elementary school students*, Phd thesis, 2018. Liaoning Normal University, Liaoning, China. Available from: http://www.cnki.net
31. Shah, D.A. and Madden, L.V., Nonparametric analysis of ordinal data in designed factorial experiments. *Phytopathology*, 2004, 94(1): 33–43. https://doi.org/10.1094/phyto.2004.94.1.33
32. Fay, M.P. and Proschan, M.A., Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics Surveys*, 2010, 4. https://doi.org/10.1214/09-ss051
33. Pande, P. and Chandrasekharan, S., Expertise as Sensorimotor Tuning: Perceptual Navigation Patterns Mark Representational Competence in Science. *Res Sci Educ*, 2022, 52(2): 725–747. https://doi.org/10.1007/s11165-020-09981-3
34. Ainley, M., Hidi, S. and Berndorff, D., Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology*, 2002, 94(3): 545–561. https://doi.org/10.1037/0022-0663.94.3.545
35. Hegarty, M. and Kozhevnikov, M., Types of visual ‐ spatial representations and mathematical problem solving. *Journal of Educational Psychology*, 1999, 91(4): 684–689. https://doi.org/10.1037/0022-0663.91.4.684
36. Greer, D.L., Crutchfield, S.A. and Woods, K.L., Cognitive Theory of Multimedia Learning, Instructional Design Principles, and Students with Learning Disabilities in Computer-based and Online Learning Environments., *Journal of Education*, 2013, 193(2): 41–50. https://doi.org/10.1177/002205741319300205
37. Mayer, R.E., The past, present, and future of the Cognitive Theory of Multimedia Learning. *Educational Psychology Review*, 2024, 36(1). https://doi.org/10.1007/s10648-023-09842-1
38. Rutten, N., Van Joolingen, W.R. and Van Der Veen, J.T., The learning effects of computer simulations in science education., *Computers & Education*, 2011, 58(1): 136–153. https://doi.org/10.1016/j.compedu.2011.07.017
39. Nisbett, R.E., The geography of thought: how Asians and Westerners think differently--and why, In *Free Press eBooks*, 2003. https://doi.org/10.26522/brocked.v13i2.52
40. Erbas, A.K., Alacaci, C. and Bulut, M., A Comparison of Mathematics Textbooks from Turkey, Singapore, and the United States of America. *Educational Sciences Theory & Practice*, 2012, 12(3): 2324–2329.
41. Sweller, J., Instructional design. *Encyclopedia of evolutionary psychological science*, 2021, 4159–4163. Cham: Springer International Publishing. https://doi.org/10.1002/0470018860.s00673
42. Duchowski, A.T., *Eye tracking Methodology: theory and practice*, In Springer eBooks, 2003. https://doi.org/10.1007/978-1-4471-3750-4
43. Rim, N.W., Choe, K.W., Scrivner, C. and Berman, M.G., Introducing Point-of-Interest as an alternative to Area-of-Interest for fixation duration analysis. *PLoS ONE*, 2021, 16(5): e0250170. https://doi.org/10.1371/journal.pone.0250170

Author's biography

Shumeng Ni is an undergraduate student at Shanghai Jiao Tong University. Her research interests include STEM education, educational psychology, and psychology linguistics.

Zhujun Jiang is a PhD candidate at the School of Education at Shanghai Jiao Tong University. Her research interests are STEM education and representation competence.

Fengkuang Chiang is a professor at the School of Education at Shanghai Jiao Tong University. His research focused on STEM education, learning space, and integration of information technology and curriculum.