International Journal of Applied Linguistics

WILEY

ORIGINAL ARTICLE OPEN ACCESS

The Relationships Between Lower- and Higher-Level Cognitive Skills and Multimodal Reading Comprehension Among Fourth-Grade Students in the Digital Age

Yaping Liu[1] | Choo Mui Cheong[2] | Xinhua Zhu[1]

[1]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China | [2]Faculty of Education, The University of Hong Kong, Hong Kong SAR, China

Correspondence: Choo Mui Cheong (cheongcm@hku.hk)

Received: 8 April 2025 | **Revised:** 31 May 2025 | **Accepted:** 10 June 2025

Funding: This study was supported by Hong Kong Quality Education Fund (Project No. 2020/0599).

Keywords: inference | linguistic and image comprehension | multimodal reading | prior knowledge | word decoding

关键词: 多模態閱讀 | 文字解碼 | 語言及圖像理解 | 推理 | 背景知識

ABSTRACT

Multimodal reading skills are essential for 21st-century students to interpret and navigate information across various modalities on the Internet and in multimedia environments. While previous studies have delved into the effects of lower- and higher-level cognitive skills on traditional reading comprehension, which involves solely written texts, relatively, few have investigated their influence in the context of multimodal reading, where texts are combined with images, audio, video, and other modalities. This study aimed to explore the relationships among lower-level cognitive skills (i.e., word decoding, listening comprehension, and image comprehension), higher-level cognitive skills (i.e., prior knowledge and inference), and multimodal reading. A total of 251 fourth grade students from Hong Kong participated in this study. Structural equation modeling (SEM) was conducted to analyze the data. The results indicated that both lower- and higher-level skills positively influence multimodal reading. Moreover, the mediating role of inference between lower-level skills and multimodal reading and the moderating role of word decoding between inference and multimodal reading were identified. This study enriches the literature on the influence of various cognitive factors in the multimodal reading process and offers significant implications for enhancing multimodal reading skills and designing effective instructional strategies.

摘要

多模態閱讀能力是21世紀學生在互聯網和多媒體環境下處理跨模態信息 (即文字、圖像、音頻及視頻等多種模態相結合) 的核心技能。以往的研究多關注低階和高階認知技能與以文字理解為主的傳統閱讀理解的關聯, 甚少對多模態閱讀與認知技能之間的關係進行探索。本研究旨在探討低階認知技能 (包括文字解碼、語言理解及圖像理解) 、高階認知技能 (包括背景知識及推理能力) 與多模態閱讀之間的關係。研究對象為251名香港小學四年級學生。透過結構方程模型分析, 本研究發現學生的低階與高階認知技能均顯著正向預測多模態閱讀表現。推理能力在低階認知技能與多模態閱讀之間扮演中介的角色, 而文字解碼能力則在推理與多模態閱讀之間起調節作用。本研究不僅豐富了多模態閱讀與不同認知因素之間關係的理論探討, 還為提升學生多模態閱讀能力 (如設計有效的教學策略) 提供了理論依據和實踐啟示。

1 | Introduction

The widespread use of the Internet and communication technologies has changed the way people construct meaning, giving rise to the concept of multimodal reading, in which meaning is constructed through multiple modes of communication, such as verbal, visual, auditory, and gestural (Jewitt and Kress 2003; The New London Group 1996). Multimodal reading in this study primarily involves text and images, as these are most common for primary school students. Compared to traditional reading comprehension, which is linear and text-based, multimodal reading requires readers to not only comprehend textual information but also interpret and integrate information from different modalities. This type of reading requires higher-level cognitive skills such as cross-modal inference and integration (Kendeou et al. 2020; Lee et al. 2021). Nowadays, multimodal reading has been incorporated into the language curricula of countries/regions such as Singapore, New Zealand (NZ), British Columbia (BC), Canada, and so on (Ministry of Education British Columbia 2022; Ministry of Education New Zealand 2025; Ministry of Education Singapore 2024), where the development of students' multimodal reading skills has become an important educational goal. At the same time, international reading literacy assessment programs, such as the Program for International Student Assessment (PISA) and the Progress in International Reading Literacy Study (PIRLS), have progressively introduced digital assessments that are rich in multimodal content and specifically target multimodal reading skills. However, researchers revealed that students encounter significant cognitive challenges when engaging with multimodal texts. For instance, they often overlook critical information presented in images (Hannus and Hyönä 1999), struggle to integrate visual and textual content deeply, and face difficulties drawing effective inferences (Cromley et al. 2013). To address these challenges and enhance students' proficiency in multimodal reading, it is essential to investigate the cognitive skills linked to multimodal reading comprehension. A detailed understanding of the specific cognitive mechanisms and processes will provide a strong theoretical basis and practical guidance for the development of effective instructional interventions and strategies.

Previous studies have examined the direct or indirect relationships between lower-level (e.g., word decoding and linguistic comprehension), higher-level cognitive skills (e.g., prior knowledge and inference), and traditional reading comprehension (Gough and Tunmer 1986; Kalantzis and Cope 2012; Tighe et al. 2023). However, little is known about how these skills synergistically contribute to multimodal reading comprehension, and the following unresolved issues hinder our initial understanding of this type of reading. First, according to the simple view of reading (SVR; Gough and Tunmer 1986), word decoding and linguistic comprehension are fundamental to reading comprehension. Critics argue that prior validations of SVR are largely based on relatively simple reading tasks and may not fully apply to cognitively complex tasks (e.g., multimodal reading; Joshi et al. 2012; Tunmer and Chapman 2012). Furthermore, multimodal reading also necessitates basic image comprehension skills, an area that lacks investigation, especially in relation to reading outcomes. Second, for higher-level skills, the critical role of prior knowledge activation and inference in traditional reading has been well-documented (e.g., Coiro and Dobler 2007; Kendeou et al. 2020). However, their contributions to multimodal reading have not been as thoroughly explored. Moreover, multimodal reading demands more complex inference skills, requiring students to infer information not only within or across texts in a single modality but also across multiple modalities and intermodalities. Third, the interaction model of reading suggests that lower- and higher-level skills interact during comprehension (Rumelhart and Ortony 1977; Stanovich 1980). Previous research indicated that higher-level skills, such as inference, play the role of mediators between lower-level skills and reading comprehension (Silva and Cain 2015), whereas lower-level skills, such as word decoding, moderate the relationship between higher-level skills and reading comprehension (Perfetti and Adlof 2012). However, these interactions in the context of multimodal reading remain unverified.

To address these research gaps, this study aims to investigate the relationships among lower-level skills (e.g., word decoding, linguistic comprehension, and image comprehension), higher-level skills (e.g., prior knowledge and inference), and multimodal reading.

2 | Literature Review

2.1 | The Theoretical Framework

Researchers have proposed several psychological models of reading to explain how readers comprehend text, extract meaning, and construct coherent mental representations during reading. These include bottom-up, top-down, and interactive models. The bottom-up model (e.g., Garnham and Oakhill 1985; Stanovich 1980) posits that information processing begins with decoding textual information, progressing incrementally to construct higher-level meaning. In this model, readers primarily rely on textual elements, such as letters, words, and syntactic structures, to interpret the content word by word. The top-down model (e.g., Davoudi 2005; Goodman 1967), on the other hand, emphasizes the influence of higher-level cognitive factors such as readers' subjective expectations, prior knowledge, experience, and inference on reading comprehension. According to this model, readers actively use their prior knowledge, schema, expectations, and contextual information to make sense of the text. The interaction model (e.g., Rumelhart and Ortony 1977) integrates both bottom-up and top-down models, viewing reading comprehension as a dynamic interaction between textual information and readers' existing knowledge structures and inference abilities. The model emphasizes the basic linguistic decoding process and the role of higher-level cognitive skills such as active inference and the invocation of prior knowledge. Modern reading research tends to support the interactive model as a more comprehensive explanation of the reading comprehension process.

Regarding multimodal reading, Mayer (2001) proposed the Cognitive Theory of Multimedia Learning (CTML), which explores how symbolic resources in multimedia can be utilized to facilitate learning. CTML is grounded in three principal psychological assumptions: the dual-channel assumption, the limited-capacity assumption, and the active-processing assumption. The dual-channel assumption posits that the human brain processes information through two independent channels: the verbal and the visual channels. The limited-capacity assumption suggests

that each channel has a finite capacity, and information overload will reduce learning effectiveness. Finally, the active-processing assumption holds that learning is an active process in which learners construct mental representations based on their prior knowledge and experience. According to CTML, the cognitive process of multimedia learning involves selecting relevant verbal and visual information, organizing this information into coherent structures, and integrating it with prior knowledge to achieve a deep understanding.

In conclusion, the models of reading elucidate the interaction between lower- and higher-level cognitive processes that readers employ to construct coherent mental representations. The CTML highlights how learners process information across various media and modalities. These theories provide a theoretical framework for understanding the cognitive process of multimodal reading and offer valuable insights into the interplay between different levels of cognitive factors. Building on this foundation, the following section delved into the collaborative relationship between lower- and higher-level cognitive skills and multimodal reading.

2.2 | Lower-Level Skills in Multimodal Reading

Word decoding and linguistic comprehension are fundamental components of reading (Gough and Tunmer 1986). In the context of multimodal reading, image comprehension is equally essential.

2.2.1 | The Chinese Writing System, Word Decoding, and Linguistic Comprehension

According to the SVR proposed by Gough and Tunmer (1986), decoding is defined as the ability to accurately and efficiently recognize words, while linguistic comprehension refers to the ability to interpret sentences and discourse using lexical knowledge. The SVR has been supported by extensive empirical research, especially in alphabetic languages such as English (García and Cain 2014; Quinn and Wagner 2018).

Previous studies highlight that the effectiveness of the SVR varies with the orthographic characteristics of the target language (e.g., Florit and Cain 2011; Joshi 2018). The SVR was originally developed for alphabetic languages like English, with relatively high phonological consistency and orthographic transparency. In contrast, Chinese is a morpho-syllabic script in which characters serve as the basic unit (e.g., DeFrancis 1984). These characters are visually and orthographically complex. Visually, Chinese characters are composed of strokes and radicals that convey phonological or semantic information, and carry more visual detail than alphabetic scripts (Li et al. 2012). Moreover, the orthography of Chinese characters is opaque, with a lack of one-to-one correspondence between phonemes and graphemes, which greatly increases the difficulty of decoding and understanding Chinese (Ho et al. 2017).

Previous studies have explored the roles of word decoding and linguistic comprehension in Chinese reading comprehension. For instance, Joshi et al. (2012) investigated the applicability of the SVR to orthographies beyond English, comparing elementary school students learning transparent Spanish, English, and opaque Chinese. They found that word decoding and linguistic comprehension explained 60% and 50% of the variance in Spanish and English reading comprehension, respectively, but only 25%–42% in Chinese. Additionally, in Spanish and English, word decoding explained more variance in the second grade, while linguistic comprehension contributed more in the fourth grade. Conversely, in Chinese, the contribution of word decoding persisted across grades for a longer period, primarily due to the slower maturation of decoding skills stemming from the orthographic complexity of Chinese. Yeung et al. (2016) followed 369 Hong Kong first graders for three years, finding that decoding and linguistic comprehension explained 36%–55% of the variance in reading comprehension at the initial stage. Structural equation modeling (SEM) confirmed both as core components of Chinese reading, with decoding exerting a stronger influence. In a recent meta-analysis, Peng et al. (2021) systematically investigated a model incorporating meta-linguistic skills, word decoding, linguistic comprehension, and Chinese reading comprehension, using data from 49,416 individuals across 210 studies. Findings indicated a moderate correlation between word decoding and linguistic comprehension, which together explained 52.7% of the variance in reading comprehension. Unlike the results of Joshi et al. (2012) and Yeung et al. (2016), they found that linguistic comprehension contributes more than word decoding. Overall, these findings suggest that while decoding and linguistic comprehension explain less variance in Chinese reading compared to alphabetic languages, they remain fundamental components of Chinese reading comprehension. The relative contributions of word decoding and linguistic comprehension in Chinese require further investigation.

Beyond that, the adaptability of SVR has been questioned, as much of the supporting evidence is based on studies employing relatively simple texts and low-inference reading tasks (Joshi et al. 2012; Tunmer and Chapman 2012). Given that multimodal reading comprehension involves higher-level cognitive processes to synthesize and integrate information across various modalities, further research is needed to examine the contributions of word decoding and linguistic comprehension in both the higher-level processes and the understanding of multimodal texts.

2.2.2 | Image Comprehension

Image comprehension is an essential skill for multimodal reading that includes recognition and meaning-making (Heinich 1999). Recognition is the ability to relate image symbols to their real-life counterparts and to understand the meaning of the image symbols. Meaning-making refers to the ability to interpret and derive meaning or value from these symbols. These cognitive functions facilitated by image comprehension are beneficial to reading comprehension (Read and Smith 1982; Sroufe 2004). For instance, Read and Smith (1982, p. 930) found that children's proficiency in understanding images enhanced their ability to comprehend text. They argued that practicing comprehension through images can "bridge the transition to print when formal reading instruction begins." Similarly, Sroufe (2004) demonstrated that integrating graphic elements with text not only facilitates reading comprehension but also strengthens oral and written communication skills.

However, researchers have observed that images convey more information than text, and their inclusion increases the complexity of the reading process, thereby raising the cognitive demands of multimodal reading (Baker and Behrens 2015). This added complexity may hinder comprehension to some extent. Therefore, basic image comprehension skills are essential for effectively understanding multimodal content. Nevertheless, further research is needed to elucidate the extent to which it contributes to multimodal reading.

2.3 | Higher-Level Skills in Multimodal Reading

In the process of reading comprehension, the activation of prior knowledge and inference are the two most common advanced skills (Coiro and Dobler 2007).

2.3.1 | Prior Knowledge

Prior knowledge is defined as knowledge that a person has acquired prior to engaging in a specific learning task (Dochy 1994). Numerous studies have shown that prior knowledge is a strong predictor of reading comprehension (e.g., Bråten et al. 2009; Florit et al. 2020; Karimi 2018). It particularly facilitates the immediate and effortless integration of information that may be implicit or missing (Ozuru et al. 2009). In a hypertext environment, researchers have demonstrated that individuals with substantial prior knowledge more effectively understand and use diverse modes of information compared to readers with low levels of prior knowledge (Coiro and Dobler 2007; Potelle and Rouet 2003). Furthermore, Kendeou et al. (2020) noted that images in multimodal reading are more likely to activate students' prior knowledge, which in turn promotes comprehension. However, some studies suggested that the role of prior knowledge in multimodal reading may be limited, as the impediments to comprehension caused by limited prior knowledge may be able to be alleviated by the supporting role of pictures. For example, ChanLin's (1998) experimental study on the use of computer images to support the understanding of scientific concepts among students with different levels of prior knowledge showed that, for those with limited prior knowledge, learning materials containing images (both static and moving) were superior to text-only in learning descriptive knowledge. However, there was no significant difference in information processing with or without images when acquiring procedural knowledge among the low prior knowledge group. This implies that deficits in prior knowledge may not substantially impede the learning of descriptive content, as the inclusion of images reduces learning complexity and eases information processing. Overall, although previous studies have explored the role of prior knowledge in multimodal learning contexts, findings remain inconsistent, and empirical support for the role of prior knowledge in enhancing multimodal reading comprehension is still lacking.

2.3.2 | Inference

Inference is defined as the "computation of implicit information" (Vonk and Noordman 2012, 447) and is widely acknowledged as a critical advanced skill necessary for reading comprehension (Kintsch 1998; McNamara and Magliano 2009; Rapp et al. 2007). The importance of inference in reading comprehension has been confirmed by a number of empirical studies (e.g., Attaprechakul 2013; Daugaard et al. 2017; Tarchi 2015). Inference is critical in multimodal reading because multimodal texts are not only richer in content, but also contain a more complex inference space. Multimodal reading generally consists of both text and images: text consists of clauses, each of which contains subjects and verbs describing actions and events (Zwaan et al. 1995), whereas images show multiple characters involved in one or more actions and events (Kendeou et al. 2020), thus providing more units of inference than text. In addition, while traditional text comprehension focuses on linguistic processing, multimodal reading requires the integration of verbal cues and visual elements, which creates conditions for more complex inference.

Several investigations have explored the relationship between inference and reading comprehension of complex forms. For example, Lodewijks (1982) observed that students with strong inference skills performed better when reading a set of passages on a related topic. Kendeou et al. (2020) developed an inferential language comprehension (iLC) framework, highlighting that inference skills are transferable across different media, significantly influencing reading comprehension. Despite these insights, empirical studies on the efficacy of inference in multimodal reading contexts remain comparatively underdeveloped.

2.4 | Relationships Among Lower- and Higher-Level Skills and Multimodal Reading

Beyond the direct relationships of both the lower- and higher-level skills with multimodal reading, the previous studies also explored the interplay between the two. These studies primarily provided the following insights that lay the foundation of the current study.

First, higher-level skills (e.g., inference) are built upon lower-level skills. Previous research identified inference acting as a mediator between lower-level skills and reading comprehension. For example, Silva and Cain (2015) assessed how lower-level comprehension skills (receptive vocabulary and grammar) supported higher-level comprehension skills (inference and story comprehension) and the role of these skills in predicting subsequent reading comprehension among 82 children aged 4–6 years. Their results supported the independent contribution of lower- and higher-level comprehension skills to reading comprehension one year later, and confirmed that the effect of lower-level skills on reading comprehension is mediated through inference skills. Similarly, Tighe et al. (2023) examined the relationships among text-based (word decoding, fluency), language-based (morphological awareness, vocabulary, linguistic comprehension), inference skills, and reading comprehension among adult readers with learning difficulties. Findings indicated that vocabulary and morphological awareness indirectly predicted reading comprehension through inference skills.

Second, lower-level skills (e.g., word decoding) can moderate the relationship between higher-level skills and reading comprehension. According to the lexical quality hypothesis (Perfetti 2007), successful comprehension depends on the degree of precision

and automatization of an individual's lexical representation. Decoding is the starting point and necessary condition for building high-quality lexical representations. In the process of text comprehension, cognitive resources are shared between decoding and higher-level skills such as inference. According to this theory, decoding could mediate the relationship between inference and reading comprehension (Perfetti and Adlof 2012). Specifically, proficient decoders require fewer cognitive resources for decoding, allowing them to allocate more resources to higher-level processes like inference, which significantly enhances reading comprehension. Conversely, insufficient decoding ability may overburden cognitive resources, dysfunctioning higher-level processing and thereby hindering comprehension. The moderating role of decoding is particularly noteworthy in Chinese reading comprehension. As Ho et al. (2003) pointed out, the orthographic complexity of Chinese characters requires stronger decoding abilities, which help reduce memory load and, in turn, facilitate deeper semantic processing during text reading.

Previous studies provide empirical support for the lexical quality hypothesis and the moderating role of decoding (Hamilton et al. 2013; Prior et al. 2014). For example, Prior et al. (2014) observed that individuals with advanced decoding skills demonstrated more automatic lexical processing, which consequently liberated cognitive resources for higher-level processing, thereby enhancing reading comprehension. Similarly, Hamilton et al. (2013) further explored the interaction between word decoding and working memory in adults, finding that word decoding significantly moderated the relationship between working memory and reading comprehension. These studies collectively underscore the critical interplay between decoding proficiency and cognitive resource allocation in the facilitation of effective reading comprehension.

Thus, for cognitively demanding and information-rich multimodal reading tasks, we hypothesized that decoding plays an important moderating role between inference and multimodal reading. However, empirical evidence to substantiate this hypothesis remains limited.

2.5 | The Present Study

In response to identified gaps within the literature, this study aimed to examine the relationship between lower- and higher-level skills and students' multimodal reading comprehension. Accordingly, the study is guided by the following research questions (RQs):

- RQ1: To what extent are lower-level skills (word decoding, linguistic comprehension, and image comprehension) and higher-level skills (prior knowledge and inference) directly associated with multimodal reading comprehension?
- RQ2: Does inference mediate the relationship between lower-level skills and multimodal reading comprehension?
- RQ3: Does word decoding moderate the relationship between inference and multimodal reading comprehension?

An SEM was constructed to address the RQs. Within SEM, a mediator refers to a variable (i.e., "inference" in this study) serving as an intermediary between the independent and the dependent variable, explaining the mechanism or process through which the independent variable relates to the dependent variable (Baron and Kenny 1986). In other words, the mediator clarifies how or why a particular relationship occurs. A moderator refers to a variable (i.e., "word decoding" in this study) that modifies the strength of the relationship between the independent and dependent variables (Baron and Kenny 1986). Specifically, a moderator can strengthen or weaken this relationship under the condition of different moderators. By comprehensively considering the roles of mediator and moderator variables, this study can better clarify how lower-level and higher-level cognitive skills relate to and interact with multimodal reading comprehension.

3 | Method

3.1 | Participants

A total of 251 fourth grade students from Hong Kong, comprising 124 girls (49.4%) and 127 boys (50.6%), voluntarily participated in the study. Their ages ranged from 9 to 12 years (M = 9.61, SD = 0.91). Three primary schools from Kowloon and the New Territories were selected through convenience sampling. Moreover, the schools were selected across different achievement levels to ensure the sample's representativeness. All participants were native Cantonese speakers. Informed consent was obtained from the students, parents or guardians, and the schools before data collection.

3.2 | Measures

This study examined the factors influencing multimodal reading performance by assessing cognitive skills such as word decoding, linguistic comprehension, image comprehension, prior knowledge, and inference.

Word decoding. Word decoding is typically measured through spelling or reading real and pseudo words aloud (Chen and Vellutino 1997). This study adopted Pan and Kang's (2003) study, which categorizes 3000 commonly used Chinese characters into six levels corresponding to grades 1 through 6 in Hong Kong primary schools. From this, 100 fourth-grade-level Chinese characters were selected, including 50 single characters and 25 two-character vocabulary, with the latter based on the list of basic Chinese words for primary students in Hong Kong (EDB 2007). Participants were instructed to read the characters aloud accurately and instantly. The test has a total score of 100 points, with one point awarded for each correctly read character. Moreover, the task demonstrated strong reliability, with a Cronbach's alpha of 0.87.

Linguistic comprehension. We employed a listening comprehension test to measure linguistic comprehension. As Hoover and Gough (1990, 157) clearly explained, the term linguistic comprehension is synonymous with "auding", defined as "listening for the purpose of comprehension". In SVR (Gough and Tunmer 1986), linguistic comprehension refers to the ability to acquire lexical information and derive sentence and discourse interpretations, which are effectively assessed through listening

comprehension tests. While reading comprehension involves similar abilities, it relies on graphic-based input through visual processing. Moreover, if linguistic comprehension refers specifically to reading comprehension, using linguistic comprehension to explain reading comprehension would introduce circular reasoning. Thus, Hoover and Gough (1990, 131) suggest that "measures of linguistic comprehension must assess the ability to understand language, for example, by assessing the ability to answer questions about the content of a listening narrative." Based on this, subsequent researchers have also commonly used listening tests to measure linguistic comprehension (e.g., Tunmer and Chapman 2007).

The listening test used in this study was adapted from the 2022 Hong Kong Territory-wide System Assessment (TSA) (HKEAA 2023) Grade 6 exam and simplified it for Grade 4 participants by reducing the format from a three-person dialogue to a two-person one. Experienced Chinese language teachers reviewed the adaptation to ensure its suitability, and teachers at participating schools confirmed that students had not previously encountered the material. The dialouge is between an uncle and a primary school student on healthy dietary choices, was recorded into a 10 minutes audio file and has two parts. The first discussing the drawbacks of sugary drinks and the second on the importance of reading food labels. 13 questions followed the dialogue, targeting three listening skills: understanding the general content, distinguishing brief comments about people and events, and identifying connections within the dialogue. Students were given an extra 10 minutes to answer the questions. The recording was produced by a local company. The test's reliability (Cronbach's alpha) was 0.66, which meets acceptable standards per Hair (2014).

Image comprehension. This study developed an image comprehension test inspired by Barbot et al.'s (2013) measurement. In their study, students were shown an image and asked three questions: a general opinion question (e.g., "Let's talk about it"), a descriptive question about image elements (e.g., "What do you notice?"), and a simple interpretive question (e.g., "What is happening in the story?"). Adopting this measurement structure, this study converted oral questions into short-answer questions in written form using traditional Chinese characters commonly used in Hong Kong. The written format was chosen to ensure standardized administration and scoring (Bachman 1990). Since image comprehension involves students' subjective perceptions and oral tests may vary with the examiner's tone, pace, or emphasis, the written test ensures identical input for all participants, minimizing potential bias and enhancing result reliability. Moreover, it is worth emphasizing that in scoring this task, we assessed only students' comprehension of the image, not their ability to write Chinese characters or their language expression skills.

Two illustrative images were selected from the Key Stage 2 (the second stage of primary education) speaking tests of the 2019 and 2023 TSA to match fourth-grade proficiency levels. The first image depicted children playing with a scooter near a waterfront, highlighting bystanders' reactions to an impending collision. The second showed a family hiking in a park, with one member experiencing heat stroke, and the group's response was emphasized. Students were required to identify image elements, discern the emotions of specific characters, and narrate the depicted scenarios. To further assess image recognition without prior knowledge bias, an "odd-one-out" task was added. Participants were asked to examine a set of pictures and identify any that did not match the characteristics of the others. The test consists of 11 items with a total score of 26 points. Cronbach's alpha for this test was 0.78.

Inference. In previous studies, the Intertextual inference verification task (IIVT) has been used to measure students' inference ability to synthesize information from multiple texts (Braasch et al. 2014; Cheong et al. 2019). In the present study, this task was adapted to evaluate inference skills using multimodal reading materials. This task included two passages on intelligence, combining text and images, with 14 items (8 valid inferences, 6 invalid inferences). Students judged each statement as "Yes" (valid) or "No" (invalid) based on the multimodal materials, which required them to integrate information from the textual and visual portions of the passage and to make inferences across modalities. The test consists of 14 items with a total score of 14 points. Scores were based on the number of correct responses. The task demonstrated strong reliability, with a Cronbach's alpha of 0.81.

Prior knowledge. To measure prior knowledge, this study designed 10 true/false items on the nature protection topic (i.e., the topic of the multimodal reading test). These items aligned with the reading task's focus on ecological principles, conservation methods, and the impact of human activities on nature, ensuring a comprehensive evaluation of students' baseline understanding of environmental conservation. The total score for the test is 10. The instrument demonstrated acceptable reliability, with a Cronbach's alpha of 0.69.

3.3 | Data Collection

The four cognitive skills tasks (excluding the prior knowledge test) were divided into two parts to prevent fatigue when administered simultaneously. These tasks were conducted within two weeks after the multimodal reading task, while the prior knowledge test was administered beforehand to avoid interference with the assessment of prior knowledge by information that students might have been exposed to and learned during the reading task.

For the word decoding task, three undergraduate students from the department of Chinese language education and research assisted in administering the test. The students were placed in a quiet classroom and read aloud pre-printed words one at a time. Research assistants evaluated the accuracy of their readings, which were audio-recorded with consent to allow for later verification.

The prior knowledge, multimodal reading comprehension, listening comprehension, image comprehension, and inference tests were administered in a standard classroom setting. The prior knowledge test lasted 10 min, the multimodal reading comprehension test 40 min, the listening comprehension test 30 min, the image comprehension test 20 min, and the inference test 30 min. After data collection, students' responses were scored. For short-answer items in the visual decoding test and

TABLE 1 | Bivariate correlations between study variables.

	1	2	3	4	5	6
1. Word decoding	1					
2. Listening comprehension	0.38**	1				
3. Image comprehension	0.48**	0.30**	1			
4. Inference	0.33**	0.30**	0.27**	1		
5. Prior knowledge	0.45**	0.23**	0.34**	0.37**	1	
6. Multimodal reading	0.50**	0.36**	0.41**	0.50**	0.40**	1
Mean	64.15	6.43	15.01	7.64	5.8	12.11
Standard deviation	20.77	1.86	3.25	1.99	1.5	6.99
Skewness	−0.90	−0.37	−0.99	−0.42	−0.66	0.49
Kurtosis	0.73	−0.07	1.47	0.01	0.16	−0.79

Note: ** $p < 0.01$.

multimodal reading comprehension, reference answers were provided for each level. These reference answers were based on students' actual responses, ensuring the scoring criteria accurately reflected their understanding and interpretations.

3.4 | Statistical Analyses

First, the confirmatory factor analysis (CFA) was conducted to validate the constructs of all instruments used. Second, a SEM model was developed using Mplus 8.0 (Muthén and Muthén 2017) with maximum likelihood estimation to examine the relationships among the study variables. To test the significance of the mediation, the study used 1,000 bootstrap samples to calculate a 95% confidence interval for the indirect relations. If the confidence interval did not contain zero, the mediation was considered significant (Preacher and Hayes 2008). For moderation, the interaction modeling method was employed by constructing an interaction term (*word decoding* × *inference*) between the independent variable (i.e., inference) and the moderating variable (i.e., word decoding). This interaction term was incorporated into the SEM for testing. Additionally, a simple slope analysis was conducted to further examine the moderating role of word decoding. Model fit was evaluated using indices such as the comparative fit index (CFI), Tucker-Lewis index (TLI), root mean squared error of approximation (RMSEA), and standard root mean square residual (SRMR). CFI and TLI values above 0.90 indicate acceptable model fits; RMSEA and SRMR values below 0.08 indicate acceptable model fits (Marsh et al. 2004).

4 | Results

4.1 | Preliminary Analysis

Table 1 presents the descriptive statistics of all the study variables and the bivariate correlations between study variables. The results showed that the absolute values of skewness (ranging from 0.37 to 0.99) and kurtosis (ranging from 0.01 to 0.79) were within acceptable ranges (|skewness| < 3 and |kurtosis| < 10), indicating that the data satisfy the assumptions of normal distribution (Kline 2023). There is a moderate correlation between the variables, ranging from 0.23 to 0.50, suggesting the possibility of further analysis.

4.2 | Relationships Among Lower-, Higher-Level Skills, and Multimodal Reading

The SEM model was conducted to explore the relationships among word decoding, linguistic comprehension, image comprehension, inference, prior knowledge, and multimodal reading comprehension. The analysis indicated a satisfactory model fit: χ^2 (5) = 25.18, CFI = 0.95, TLI = 0.89, RMSEA = 0.08, and SRMR = 0.07.

In response to RQ1, the independent contributions of lower- and higher-level skills were examined. As illustrated in Figure 1, both lower- and higher-level skills were significantly associated with multimodal reading comprehension. Specifically, multimodal reading comprehension was positively predicted by word decoding ($\beta = 0.28$, $p < 0.001$), linguistic comprehension ($\beta = 0.20$, $p < 0.001$), image comprehension ($\beta = 0.10$, $p < 0.05$), inference ($\beta = 0.27$, $p < 0.001$), and prior knowledge ($\beta = 0.12$, $p < 0.01$). Furthermore, the findings also showed that lower-level skills contribute to higher-level skills. In particular, word decoding ($\beta = 0.19$, $p < 0.01$), linguistic comprehension ($\beta = 0.17$, $p < 0.01$), and image comprehension ($\beta = 0.15$, $p < 0.05$) had a positive effect on inference.

To answer RQ2, the mediating role of inference was further examined. It was found that inference acts as a mediator between lower-level skills and multimodal reading comprehension (see Table 2). Specifically, word decoding ($\beta = 0.05$, 95% CI = [0.01, 0.10], $p < 0.05$), linguistic comprehension ($\beta = 0.05$, 95% CI = [0.01, 0.08], $p < 0.05$), and image comprehension ($\beta = 0.04$, 95% CI = [0.004, 0.08], $p < 0.05$) were indirectly associated with multimodal reading comprehension via inference.

In response to RQ3, the moderating role of word decoding was examined. Results showed that word decoding positively moderated the relationship between inference and multimodal

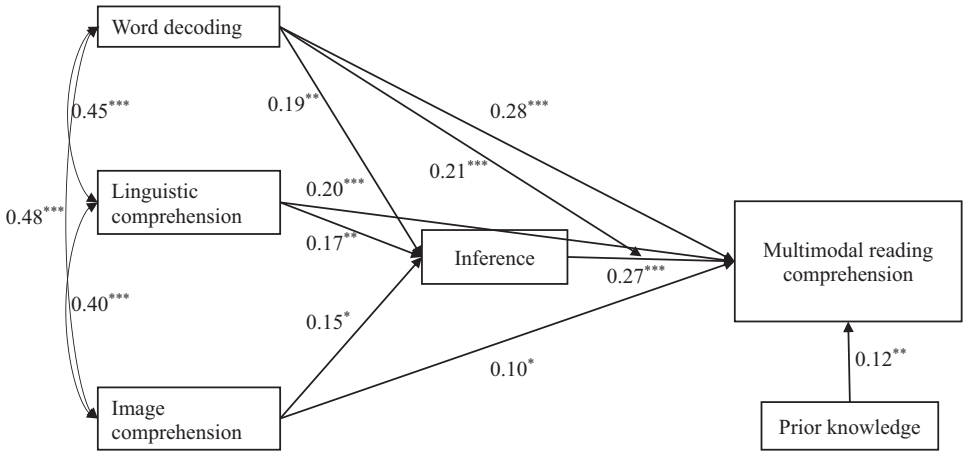

FIGURE 1 | Standardization coefficients of the model. Note: ***$p < 0.001$, **$p < 0.01$, and *$p < 0.05$.

TABLE 2 | The potential pathways linking cognitive skills to multimodal reading comprehension.

Path	Indirect effect (β)	95% CI	p
Path A. Word decoding → inference → multimodal reading	0.05	0.01, 0.10	$P < 0.05$
Path B. Linguistic comprehension → inference → multimodal reading	0.05	0.01, 0.08	$P < 0.05$
Path C. Image comprehension → inference → multimodal reading	0.04	0.004, 0.08	$P < 0.05$

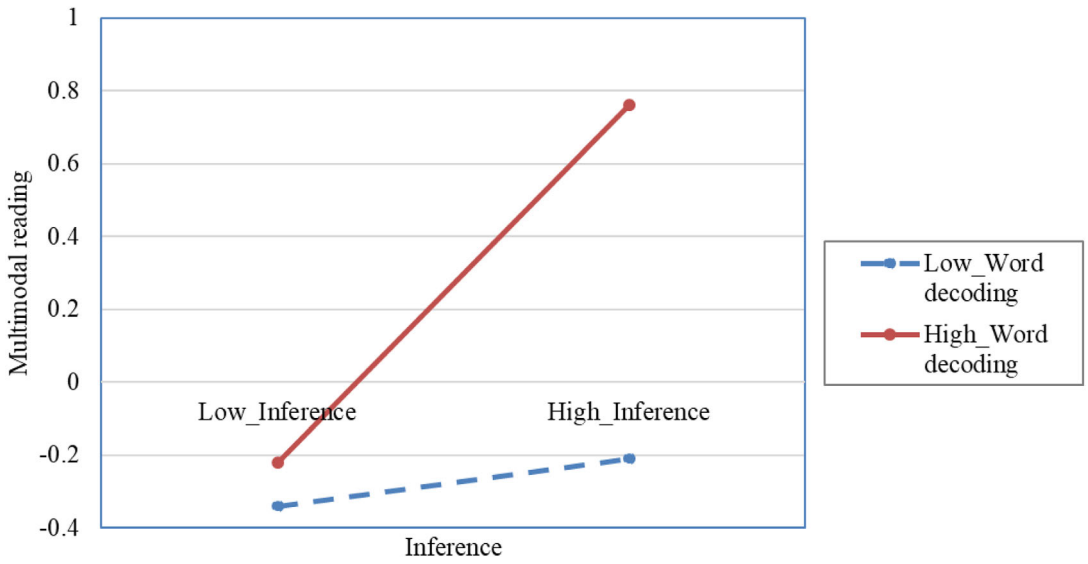

FIGURE 2 | The moderating role of word decoding on the relationship between inference and multimodal reading. [Color figure can be viewed at wileyonlinelibrary.com]

reading comprehension ($\beta = 0.21$, $p < 0.05$). To further examine this moderating effect, a simple slope analysis was conducted. Figure 2 illustrates the moderating role of word decoding by plotting the values of multimodal reading performance corresponding to word decoding and inference at plus or minus one standard deviation from their means. It can be seen that inference had a stronger positive association with multimodal reading for students with higher word decoding skills, while this association was weaker for those with lower word decoding skills.

5 | Discussion

This study aimed to examine the relationships between lower- and higher-level cognitive skills and multimodal reading comprehension. The results highlighted the essential role of lower-level skills (i.e., word decoding, linguistic comprehension, and image comprehension), while also underscoring the significance of higher-level skills like inference and prior knowledge. Additionally, lower-level skills were found to be indirectly associated with multimodal reading comprehension through inference.

Furthermore, word decoding positively moderated the association between inference and multimodal reading performance.

5.1 | The Relationships Between Word Decoding, Linguistic Comprehension, Image Comprehension, and Multimodal Reading

The findings of this study highlight the critical role of lower-level cognitive skills in facilitating multimodal reading comprehension among students in this age group. These results align with the bottom-up model and emphasize the importance of decoding and comprehension based on the text itself (Goodman 1967). Although multimodal reading materials incorporate diverse modalities, text remains the primary medium for conveying complex ideas and detailed information. The findings supported that word decoding is essential for Chinese reading comprehension, aligning with prior research on the SVR in Chinese (Joshi et al. 2012; Yeung et al. 2016). As noted, Chinese has a complex orthography with an opaque glyph-phonology relationship, increasing cognitive load during decoding (Ho et al. 2017). Strong decoding skills enable quick recognition of glyphs, phonology, and semantics, improving reading accuracy and efficiency (Quinn and Wagner 2018). In cognitively demanding tasks like multimodal reading, proficient decoding not only supports deep semantic integration but also frees cognitive resources for intermodal correlation and higher-level processing. Moreover, word decoding helps clarify and reinforce content presented through other modalities (Levin et al. 1987), promoting semantic matching and integration across various sources and ultimately enhancing overall reading effectiveness.

The findings supported that linguistic comprehension is a core skill for Chinese reading comprehension, consistent with most prior studies on the SVR in Chinese (e.g., Ho et al. 2017; Joshi et al. 2012). This study further highlights its effectiveness in cognitively complex reading tasks, addressing concerns that the SVR applies only to simple or low-inference reading tasks. In multimodal reading contexts, linguistic comprehension facilitates the rapid extraction of semantic information from the text (Hoover and Gough 1990), providing a semantic framework for interpreting information from other modalities. For instance, in scientific texts, textual descriptions provide background knowledge and conceptual frameworks that enhance the understanding of accompanying diagrams or illustrations. Moreover, linguistic comprehension is indispensable for synthesizing multimodal information to achieve a cohesive understanding. For example, when reading a scientific text with a schematic diagram of an experiment, linguistic comprehension helps align details of the diagram with the text, fostering a comprehensive grasp of the process. Furthermore, multimodal reading often involves incomplete information across modalities, requiring readers to infer and fill in gaps to construct meaning. Linguistic comprehension supports this process by extracting implicit information from the text and integrating it with other modalities through the activation of background knowledge and contextual clues (Pan and Lin 2023).

In addition, this study found that word decoding significantly contributes to Chinese reading comprehension, even among fourth graders, and this relationship is greater than linguistic comprehension. This contrasts with findings in transparent orthographies (e.g., English, Greek, and Spanish), where decoding dominates until second grade, after which linguistic comprehension becomes more stable and influential. These differences may stem from variations in orthographic transparency (Florit and Cain 2011; Joshi et al. 2012). In Chinese, the opaque orthography increases the cognitive load of word decoding compared to alphabetic writing systems, and the development of automated word decoding skills takes longer. Consequently, word decoding remains a critical component of reading comprehension for a longer developmental period in Chinese readers than in readers of alphabetic languages (Florit and Cain 2011; Yeung et al. 2016).

Our findings also underscore the contribution of image comprehension to multimodal reading. Specifically, effective image comprehension can improve recall of reading content, provide a structural scaffold for understanding textual material, support comprehension monitoring, and facilitate higher-level cognitive processes (Carney and Levin 2001; Rubman and Salatas Waters 2000). Furthermore, it is worth pointing out that the contribution of image comprehension to multimodal reading is relatively minor compared to word decoding and linguistic comprehension. One possible explanation for this is that students in the early stages of engaging with multimodal reading materials tend to rely more heavily on textual information rather than visual cues (Schmidt-Weigand et al. 2010). Second, this may be influenced by the test design. Although a certain percentage of items involved the examination of image content, more than half relied on textual comprehension. Future research could focus on task design incorporate additional items focused on image content to fully assess the contribution of image comprehension to multimodal reading.

5.2 | The Associations Between Prior Knowledge, Inference, and Multimodal Reading

The findings of this study supported the role of prior knowledge and inference in multimodal reading. According to the top-down model and CTML, prior knowledge is essential for effective comprehension and learning from multimodal reading material (Garnham and Oakhill 1985; Mayer 2001). Meaningful learning occurs when individuals successfully integrate new information from text and images with their prior knowledge to construct coherent mental representations (Richter et al. 2016; Unsworth et al. 2019). Additionally, McNamara et al. (1996) demonstrated that prior knowledge activation facilitates comprehension of discontinuous texts by enabling readers to derive meaning from content that lacks coherence. In such cases, readers rely on their prior knowledge to bridge gaps in the text, actively establishing relationships that are not explicitly presented. Consequently, prior knowledge is indispensable for understanding discontinuous multimodal reading materials.

This study highlights the critical role of inference in multimodal reading. Previous research has extensively validated the importance of inference in traditional reading contexts (Garnham and Oakhill 2014; McNamara and Magliano 2009). The present study further provides empirical evidence supporting inference as a crucial component of multimodal reading. Specifically, when students encounter textual content alongside static or dynamic visual narratives, they need to actively integrate and map textual

and visual information, generating effective inferences throughout the comprehension process (McNamara and Magliano 2009). This inferential process is crucial for constructing coherent cross-modal mental representation. Furthermore, inferential skills enable readers to assess the consistency and accuracy of presented information and to derive deeper, implicit meanings from multimodal materials.

5.3 | The Relationships Between Lower- and Higher-Level Skills and Multimodal Reading Comprehension

Beyond the direct impact of both lower- and higher-level cognitive skills on multimodal reading, more complicated influencing mechanisms exist between them, such as inference serving as a mediator and decoding as a moderator. This is consistent with the interactive model of reading, which views reading comprehension as the result of an interaction between top-down and bottom-up processing (Rumelhart and Ortony 1977). In this process, readers must not only fully utilize the information provided by the text but also rely on their prior knowledge and engage in complex inference.

In particular, the findings indicated that word decoding, linguistic comprehension, and image comprehension are indirectly associated with multimodal reading through inference. Previous research has demonstrated that lower-level cognitive skills enable readers to generate coherent inferences, thereby enhancing comprehension (Tighe et al. 2023). This study further validated the mediating role of inferences in a cognitively demanding multimodal task. According to the iLC framework, inference can be developed in a multimedia environment using multimodal information (Kendeou et al. 2020). The understanding of multimodal information not only depends on the surface-level content of different modalities but also requires readers to actively infer meaning based on visual cues and their own prior knowledge during the reading process, so as to complete the implicit information and establish the overall context. As described by Magliano et al. (2013), the lower-level skills represent the "front-end" processes of comprehension, whereas inference is a "back-end" process that enables readers to construct mental representations across diverse media and modalities.

Additionally, we found that word decoding positively moderates the relationship between inference and multimodal reading. In other words, for students with higher decoding ability, inference plays a greater role in multimodal reading comprehension. Conversely, for students with lower decoding ability, the contribution of inference is relatively smaller. These results provide empirical support for the lexical quality hypothesis, which posits that word decoding and higher-level cognitive skills, such as inference, share cognitive resources during reading. With limited cognitive resources, students with proficient decoding abilities consume fewer resources for word recognition, thereby freeing up more cognitive resources for higher-level processing, such as inference, ultimately enhancing reading performance (Perfetti and Adlof 2012). Particularly in Chinese, a language with an opaque orthography, the complex correspondence between orthography, phonology, and semantics demands significant cognitive resources. Consequently, automated word recognition becomes especially critical (Cain and Oakhill 2006). This study further highlights the important role of proficient decoding in reallocating limited cognitive resources to support advanced processing in multimodal reading contexts. In such contexts, the richness and complexity of information place higher cognitive demands, making it challenging for novice readers with insufficient decoding skills to comprehend and infer information across modalities. Conversely, as decoding becomes automated, readers can more effectively leverage their ability to integrate and infer visual and verbal information (Kirby 1993).

Overall, this study further substantiates the interaction between lower- and higher-level skills in a multimodal context, contributing to a more comprehensive and in-depth understanding of multimodal reading.

6 | Conclusions

Overall, the present study examined the mechanisms by which lower- and higher-level cognitive skills are associated with multimodal reading comprehension, highlighting three key findings: (1) word decoding, linguistic comprehension, image comprehension, inference, and prior knowledge were positively associated with students' multimodal reading performance; (2) inference mediated the relationship between lower-level skills on multimodal reading comprehension; and (3) word decoding positively moderated the association between inference and multimodal reading comprehension.

That being so, this study is not without limitations. First, a major limitation of this study is its focus on only two modalities, text and image, representing a relatively narrow definition of multimodality. While these are common in reading comprehension, they exclude other prevalent modalities in today's digital environment, such as video, audio, and interactive elements, limiting the generalizability of the findings. Future research should expand the scope to include more modalities such as video, audio, and interactive media to better reflect contemporary multimodal literacy practices. Second, the relatively low reliability coefficients for the measures of listening comprehension (0.66) and prior knowledge (0.69) present another limitation. While these values are within acceptable thresholds for research purposes (Hair 2014), they may introduce some degree of measurement error. Future research could consider adopting more robust measures or refining existing instruments to improve reliability. Third, this study used a cross-sectional design to explore the antecedents of multimodal reading performance, which can only reveal the relationship between variables at a given point in time, but does not allow for any causal inferences to be drawn (Spector 2019). Future research could use longitudinal studies or experimental designs to investigate the causal relationships between the study variables. Finally, the sample used in this study was Chinese learners in grade 4 of elementary school in Hong Kong, and the sample specificity limits the generalizability of the findings. On the one hand, learners' cognitive and reading abilities vary significantly with age and educational level (Chall 1983), and thus the relative contributions of different levels of cognitive skills may differ across grade levels (Quinn and Wagner 2018). On the other hand, orthographic transparency influences the relative contribution of decoding and linguistic comprehension

to Chinese multimodal reading comprehension (Florit and Cain 2011; Joshi 2018). Future studies could include participants from different age groups, educational levels, and languages with varying orthographic transparency to validate and enhance the representativeness of the findings.

Our study has significant theoretical and practical implications. Theoretically, it provides a comprehensive examination of the relationships between lower- and higher-level cognitive skills and multimodal reading comprehension, deepening our understanding of the hierarchy and interplay of cognitive processes involved in multimodal comprehension. Furthermore, this study supports and extends the lexical quality hypothesis by applying it to multimodal reading contexts. In practice, this study reveals the mechanisms underlying the association between lower- and higher-level cognitive skills and multimodal reading, providing valuable insights for developing and enhancing students' multimodal reading abilities. Specifically, teachers should conduct targeted training in their instruction, beginning with the reinforcement of lower-level skills such as word decoding, linguistic comprehension, and image comprehension to establish a solid basis for information acquisition. Secondly, teachers can utilize multimodal texts and digital media resources to design integrated reading activities that combine text and images, such as group discussions, case analyses, and inference exercises, thereby enhancing students' advanced cognitive skills and inference abilities. Furthermore, it is important to focus on the effective integration of lower- and higher-level skills, enabling students to gradually develop deeper information integration and critical thinking while consolidating their basic skills (Zhu et al. 2020), thus achieving a comprehensive improvement in multimodal reading.

Acknowledgments

We acknowledge the assistance of ChatGPT-4 for providing proofreading support and affirm that the research content is original.

Ethics Statement

This study was approved by Human Research Ethics Committee, The University of Hong Kong (HREC Reference Number: EA210591).

Conflicts of Interest

All authors declare that no conflicts of interest.

Data Availability Statement

Research data are not shared.

Peer Review

The peer review history for this article is available at https://publons.com/publon/10.1111/ijal.12794.

References

Attaprechakul, D. 2013. "Inference Strategies to Improve Reading Comprehension of Challenging Texts." *English Language Teaching* 6, no. 3: 82–91. https://doi.org/10.5539/elt.v6n3p82.

Bachman, L. F. 1990. *Fundamental Considerations in Language Testing*. Oxford University Press.

Baker, M., and C. F. Behrens. 2015. "Prostate Displacement During Transabdominal Ultrasound Image-Guided Radiotherapy Assessed by Real-Time Four-Dimensional Transperineal Monitoring." *Acta Oncologica* 54, no. 9: 1508–1514. https://doi.org/10.3109/0284186X.2015.1061208.

Barbot, B., J. Randi, M. Tan, C. Levenson, L. Friedlaender, and E. L. Grigorenko. 2013. "From Perception to Creative Writing: A Multi-Method Pilot Study of a Visual Literacy Instructional Approach." *Learning and Individual Differences* 28: 167–176. https://doi.org/10.1016/j.lindif.2012.09.003.

Baron, R. M., and D. A. Kenny. 1986. "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51, no. 6: 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173.

Braasch, J. L. G., I. Bråten, H. I. Strømsø, and Ø. Anmarkrud. 2014. "Incremental Theories of Intelligence Predict Multiple Document Comprehension." *Learning and Individual Differences* 31: 11–20. https://doi.org/10.1016/j.lindif.2013.12.012.

Bråten, I., H. I. Strømsø, and M. A. Britt. 2009. "Trust Matters: Examining the Role of Source Evaluation in Students' Construction of Meaning Within and Across Multiple Texts." *Reading Research Quarterly* 44, no. 1: 6–28. https://doi.org/10.1598/RRQ.44.1.1.

Cain, K., and J. Oakhill. 2006. "Profiles of Children With Specific Reading Comprehension Difficulties." *British Journal of Educational Psychology* 76, no. 4: 683–696. https://doi.org/10.1348/000709905x67610.

Carney, R. N., and J. R. Levin. 2001. "Pictorial Illustrations Still Improve Students' Learning From Text." *Educational Psychology Review* 13, no. 1: 5–26.

Chall, J. S. 1983. *Stages of Reading Development*. McGraw-Hill.

ChanLin, L. 1998. "Animation to Teach Students of Different Knowledge Levels." *Journal of Instructional Psychology* 25, no. 3: 166.

Chen, R. S., and F. R. Vellutino. 1997. "Prediction of Reading Ability: A Cross-Validation Study of the Simple View of Reading." *Journal of Literacy Research* 29, no. 1: 1–24. https://doi.org/10.1080/10862969709547947.

Cheong, C. M., X. Zhu, G. Y. Li, and H. Wen. 2019. "Effects of Intertextual Processing on L2 Integrated Writing." *Journal of Second Language Writing* 44: 63–75. https://doi.org/10.1016/j.jslw.2019.03.004.

Coiro, J., and E. Dobler. 2007. "Exploring the Online Reading Comprehension Strategies Used by Sixth-Grade Skilled Readers to Search for and Locate Information on the Internet." *Reading Research Quarterly* 42, no. 2: 214–257. https://doi.org/10.1598/RRQ.42.2.2.

Cromley, J. G., B. W. Bergey, S. Fitzhugh, et al. 2013. "Effects of Three Diagram Instruction Methods on Transfer of Diagram Comprehension Skills: The Critical Role of Inference While Learning." *Learning and Instruction* 26: 45–58. https://doi.org/10.1016/j.learninstruc.2013.01.003.

Daugaard, H. T., K. Cain, and C. Elbro. 2017. "From Words to Text: Inference Making Mediates the Role of Vocabulary in Children's Reading Comprehension." *Reading and Writing* 30, no. 8: 1773–1788. https://doi.org/10.1007/s11145-017-9752-2.

Davoudi, S. 2005. "Understanding Territorial Cohesion." *Planning Practice and Research* 20, no. 4: 433–441. https://doi.org/10.1080/02697450600767926.

DeFrancis, J. 1984. "Digraphia." *Word* 35, no. 1: 59–66. https://doi.org/10.1080/00437956.1984.11435748.

Dochy, F. 1994. "Prior Knowledge and Learning." In *International Encyclopedia of Education*, 4698–4702. Pergamon Press.

Education Bureau. 2007. *Lexical Lists for Chinese Learning in Hong Kong*. Education Bureau.

Florit, E., and K. Cain. 2011. "The Simple View of Reading: Is It Valid for Different Types of Alphabetic Orthographies?" *Educational Psychology Review* 23: 553–576. https://doi.org/10.1007/s10648-011-9175-6.

Florit, E., P. De Carli, G. Giunti, and L. Mason. 2020. "Advanced Theory of Mind Uniquely Contributes to Children's Multiple-Text Comprehension."

Journal of Experimental Child Psychology 189: 104708. https://doi.org/10.1016/j.jecp.2019.104708.

García, J. R., and K. Cain. 2014. "Decoding and Reading Comprehension: A Meta-Analysis to Identify Which Reader and Assessment Characteristics Influence the Strength of the Relationship in English." *Review of Educational Research* 84: 74–111. https://doi.org/10.3102/0034654313499616.

Garnham, A., and J. Oakhill. 1985. "On-Line Resolution of Anaphoric Pronouns: Effects of Inference Making and Verb Semantics." *British Journal of Psychology* 76, no. 3: 385–393. https://doi.org/10.1111/j.2044-8295.1985.tb01961.x.

Garnham, A., and J. Oakhill. 2014. "The Mental Models Theory of Language Comprehension." In *Models of Understanding Text*, 313–339. Psychology Press.

Goodman, K. S. 1967. "Reading: A Psycholinguistic Guessing Game." *Literacy Research and Instruction* 6: 126–135.

Gough, P. B., and W. E. Tunmer. 1986. "Decoding, Reading, and Reading Disability." *Remedial and Special Education* 7, no. 1: 6–10. https://doi.org/10.1177/074193258600700104.

Hair, J. 2014. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. SAGE Publications.

Hamilton, S. T., E. M. Freed, and D. L. Long. 2013. "Modeling Reader and Text Interactions During Narrative Comprehension: A Test of the Lexical Quality Hypothesis." *Discourse Processes* 50, no. 2: 139–163. https://doi.org/10.1080/0163853X.2012.742001.

Hannus, M., and J. Hyönä. 1999. "Utilization of Illustrations During Learning of Science Textbook Passages Among Low- and High-Ability Children." *Contemporary Educational Psychology* 24, no. 2: 95–123. https://doi.org/10.1006/ceps.1998.0987.

Heinich, N. 1999. *Art Contemporain Et Fabrication De L'inauthentique*. Association Terrain.

Ho, C. S. H., M. Zheng, C. McBride, L. S. J. Hsu, M. M. Waye, and J. C. Y. Kwok. 2017. "Examining an Extended Simple View of Reading in Chinese: The Role of Naming Efficiency for Reading Comprehension." *Contemporary Educational Psychology* 51: 293–302. https://doi.org/10.1016/j.cedpsych.2017.08.009.

Ho, C. S. H., P. W. Y. Yau, and A. Au. 2003. "Development of Orthographic Knowledge and its Relationship With Reading and Spelling Among Chinese Kindergarten and Primary School Children." *Reading Development in Chinese Children* 51–71.

Hong Kong Examinations and Assessment Authority. 2023. "Gainful Use of TSA 2022 Materials - Sub-Papers of Individual Subjects." *Hong Kong Examinations and Assessment Authority*. https://www.bca.hkeaa.edu.hk/web/TSA/zh/2022priPaper/PriIndex.html.

Hoover, W. A., and P. B. Gough. 1990. "The Simple View of Reading." *Reading and Writing* 2, no. 2: 127–160. https://doi.org/10.1007/BF00401799.

Jewitt, C., and G. R. Kress. 2003. *Multimodal Literacy*. Lang New York.

Joshi, R. M. 2018. "Simple View of Reading (SVR) in Different Orthographies: Seeing the Forest With the Trees." *Reading and Dyslexia: From Basic Functions to Higher Order Cognition* 16: 71–80. https://doi.org/10.1007/978-3-319-90805-2_4.

Joshi, R. M., S. Tao, P. G. Aaron, and B. Quiroz. 2012. "Cognitive Component of Componential Model of Reading Applied to Different Orthographies." *Journal of Learning Disabilities* 45, no. 5: 480–486. https://doi.org/10.1177/0022219411432690.

Kalantzis, M., and B. Cope. 2012. *New Learning: Elements of a Science of Education*. Cambridge University Press.

Karimi, M. N. 2018. "The Mediated/Unmediated Contributions of Language Proficiency and Prior Knowledge to L2 Multiple-Texts Comprehension: A Structural Equation Modelling Analysis." *Applied Linguistics* 39, no. 6: 912–932. https://doi.org/10.1093/applin/amw059.

Kendeou, P., K. L. McMaster, R. Butterfuss, J. Kim, B. Bresina, and K. Wagner. 2020. "The Inferential Language Comprehension (iLC) Framework: Supporting Children's Comprehension of Visual Narratives." *Topics in Cognitive Science* 12, no. 1: 256–273. https://doi.org/10.1111/tops.12457.

Kintsch, W. 1998. *Comprehension: A Paradigm for Cognition*. Cambridge University Press.

Kirby, J. R. 1993. "Collaborative and Competitive Effects of Verbal and Spatial Processes." *Learning and Instruction* 3, no. 3: 201–214. https://doi.org/10.1016/0959-4752(93)90004-J.

Kline, R. B. 2023. *Principles and Practice of Structural Equation Modeling*. Guilford Publications.

Lee, S. Y., Y. H. G. Lo, and T. C. Chin. 2021. "Practicing Multiliteracies to Enhance EFL Learners' Meaning Making Process and Language Development: A Multimodal Problem-Based Approach." *Computer Assisted Language Learning* 34, no. 1-2: 66–91. https://doi.org/10.1080/09588221.2019.1614959.

Levin, R. C., A. K. Klevorick, R. R. Nelson, S. G. Winter, R. Gilbert, and Z. Griliches. 1987. "Appropriating the Returns From Industrial Research and Development." *Brookings Papers on Economic Activity* 1987, no. 3: 783. https://doi.org/10.2307/2534454.

Li, H., H. Shu, C. McBride-Chang, H. Liu, and H. Peng. 2012. "Chinese Children's Character Recognition: Visuo-Orthographic, Phonological Processing and Morphological Skills." *Journal of Research in Reading* 35, no. 3: 287–307. https://doi.org/10.1111/j.1467-9817.2010.01460.x.

Lodewijks, H. G. L. C. 1982. "Self-Regulated Versus Teacher-Provided Sequencing of Information in Learning from Text." *Advances in Psychology* 8: 509–520. https://doi.org/10.1016/S0166-4115(08)62715-6.

Magliano, J. P., L. C. Loschky, J. A. Clinton, and A. M. Larson. 2013. "Is Reading the Same as Viewing." In *Unraveling the Behavioral, Neurobiological and Genetic Components of Reading Comprehension*, 78–90. Paul H. Brookes Publishing Co.

Marsh, H. W., Z. Wen, and K. T. Hau. 2004. "Structural Equation Models of Latent Interactions: Evaluation of Alternative Estimation Strategies and Indicator Construction." *Psychological Methods* 9, no. 3: 275–300. https://doi.org/10.1037/1082-989X.9.3.275.

Mayer, R. E. 2001. *The Cambridge Handbook of Multimedia Learning*. Cambridge University Press.

McNamara, D. S., E. Kintsch, N. B. Songer, and W. Kintsch. 1996. "Are Good Texts Always BetterZ." *Cognition and Instruction* 14, no. 1: 1–43. https://doi.org/10.1207/s1532690xci1401_1.

McNamara, D. S., and J. Magliano. 2009. "Chapter 9 toward a Comprehensive Model of Comprehension." In *Psychology of Learning and Motivation* 51: 297–384. https://doi.org/10.1016/S0079-7421(09)51009-2.

Ministry of Education British Columbia. 2022. *English Language Arts*. British Columbia's Curriculum. https://curriculum.gov.bc.ca/curriculum/english-language-arts.

Ministry of Education New Zealand. 20 March 2025. "New Zealand Curriculum—English Phase 2 – Years 46." *Te Poutāhū: Curriculum Centre, Ministry of Education*. https://newzealandcurriculum.tahurangi.education.govt.nz/nzc—english-phase-2/5637238346.p.

Muthén, L. K., and B. Muthén. 2017. *Mplus User's Guide*. Muthén & Muthén.

Ozuru, Y., K. Dempsey, and D. S. McNamara. 2009. "Prior Knowledge, Reading Skill, and Text Cohesion in the Comprehension of Science Texts." *Learning and Instruction* 19, no. 3: 228–242. https://doi.org/10.1016/j.learninstruc.2008.04.003.

Pan, D. J., and D. Lin. 2023. "Cognitive-Linguistic Skills Explain Chinese Reading Comprehension Within and Beyond the Simple View of Reading in Hong Kong Kindergarteners." *Language Learning* 73, no. 1: 126–160. https://doi.org/10.1111/lang.12515.

Pan, H., and B. Kang. 2003. *A Study of the Chinese Characters Recommended for the Subject of Chinese Language in Primary School*. Hong Kong Baptist University Language Centre.

Peng, P., K. Lee, J. Luo, S. Li, R. M. Joshi, and S. Tao. 2021. "Simple View of Reading in Chinese: A One-Stage Meta-Analytic Structural Equation Modeling." *Review of Educational Research* 91, no. 1: 3–33. https://doi.org/10.3102/0034654320964198.

Perfetti, C. 2007. "Reading Ability: Lexical Quality to Comprehension." *Scientific Studies of Reading* 11, no. 4: 357–383. https://doi.org/10.1080/10888430701530730.

Perfetti, C., and S. M. Adlof. 2012. "Reading Comprehension: A Conceptual Framework From Word Meaning to Text Meaning." *Measuring Up: Advances in How We Assess Reading Ability* 1: 3–20.

Potelle, H., and J.-F. Rouet. 2003. "Effects of Content Representation and Readers' Prior Knowledge on the Comprehension of Hypertext." *International Journal of Human-Computer Studies* 58, no. 3: 327–345. https://doi.org/10.1016/S1071-5819(03)00016-8.

Preacher, K. J., and A. F. Hayes. 2008. "Asymptotic and Resampling Strategies for Assessing and Comparing Indirect Effects in Multiple Mediator Models." *Behavior Research Methods* 40, no. 3: 879–891. https://doi.org/10.3758/BRM.40.3.879.

Prior, A., A. Goldina, M. Shany, E. Geva, and T. Katzir. 2014. "Lexical Inference in L2: Predictive Roles of Vocabulary Knowledge and Reading Skill Beyond Reading Comprehension." *Reading and Writing* 27, no. 8: 1467–1484. https://doi.org/10.1007/s11145-014-9501-8.

Quinn, J. M., and R. K. Wagner. 2018. "Using Meta-Analytic Structural Equation Modeling to Study Developmental Change in Relations Between Language and Literacy." *Child Development* 89, no. 6: 1956–1969. https://doi.org/10.1111/cdev.13049.

Rapp, D. N., P. v. d. Broek, K. L. McMaster, P. Kendeou, and C. A. Espin. 2007. "Higher-Level Comprehension Processes in Struggling Readers: A Perspective for Research and Intervention." *Scientific Studies of Reading* 11, no. 4: 289–312. https://doi.org/10.1080/10888430701530417.

Read, D., and H. M. Smith. 1982. "Teaching Visual Literacy Through Wordless Picture Books." *The Reading Teacher* 35, no. 8: 928–933.

Richter, J., K. Scheiter, and A. Eitel. 2016. "Signaling Text-Picture Relations in Multimedia Learning: A Comprehensive Meta-Analysis." *Educational Research Review* 17: 19–36. https://doi.org/10.1016/j.edurev.2015.12.003.

Rubman, C. N., and H. Salatas Waters. 2000. "A, B Seeing: The Role of Constructive Processes in Children's Comprehension Monitoring." *Journal of Educational Psychology* 92, no. 3: 503–514. https://doi.org/10.1037/0022-0663.92.3.503.

Rumelhart, D. E., and A. Ortony. 1977. "The Representation of Knowledge in Memory." In *Schooling and the Acquisition of Knowledge*, edited by R.C. Anderson R.J. Spiro, and W.E. Montague, 99–135. Erlbaum.

Schmidt-Weigand, F., A. Kohnert, and U. Glowalla. 2010. "A Closer Look at Split Visual Attention in System- and Self-Paced Instruction in Multimedia Learning." *Learning and Instruction* 20, no. 2: 100–110. https://doi.org/10.1016/j.learninstruc.2009.02.011.

Silva, M., and K. Cain. 2015. "The Relations Between Lower and Higher Level Comprehension Skills and Their Role in Prediction of Early Reading Comprehension." *Journal of Educational Psychology* 107, no. 2: 321–331. https://doi.org/10.1037/a0037769.

Singapore, M. o. E. 2024. *Chinese Language Syllabus Primary. Curriculum Planning and Development Division*. Ministry of Education.

Spector, P. E. 2019. "Do Not Cross Me: Optimizing the Use of Cross-Sectional Designs." *Journal of Business and Psychology* 34, no. 2: 125–137. https://doi.org/10.1007/s10869-018-09613-8.

Sroufe, G. E. 2004. *The Arts and Education: New Opportunities for Research*. Arts Education Partnership.

Stanovich, K. E. 1980. "Toward an Interactive-Compensatory Model of Individual Differences in the Development of Reading Fluency." *Reading Research Quarterly* 16, no. 1: 32. https://doi.org/10.2307/747348.

Tarchi, C. 2015. "Fostering Reading Comprehension of Expository Texts Through the Activation of Readers' Prior Knowledge and Inference-Making Skills." *International Journal of Educational Research* 72: 80–88. https://doi.org/10.1016/j.ijer.2015.04.013.

The New London Group. 1996. "A Pedagogy of Multiliteracies: Designing Social Futures." *Harvard Educational Review* 66, no. 1: 60–93. https://doi.org/10.17763/haer.66.1.17370n67v22j160u.

Tighe, E. L., G. Kaldes, and D. S. McNamara. 2023. "The Role of Inferencing in Struggling Adult Readers' Comprehension of Different Texts: A Mediation Analysis." *Learning and Individual Differences* 102: 102268. https://doi.org/10.1016/j.lindif.2023.102268.

Tunmer, W. E., and J. W. Chapman. 2007. "Language-Related Differences Between Discrepancy-Defined and Non-Discrepancy-Defined Poor Readers: A Longitudinal Study of Dyslexia in New Zealand." *Dyslexia: An International Journal of Research and Practice* 13, no. 1: 42–66. https://doi.org/10.1002/dys.327.

Tunmer, W. E., and J. W. Chapman. 2012. "The Simple View of Reading Redux: Vocabulary Knowledge and the Independent Components Hypothesis." *Journal of Learning Disabilities* 45, no. 5: 453–466. https://doi.org/10.1177/0022219411432685.

Unsworth, R. K. F., L. M. Nordlund, and L. C. Cullen-Unsworth. 2019. "Seagrass Meadows Support Global Fisheries Production." *Conservation Letters* 12, no. 1: e12566. https://doi.org/10.1111/conl.12566.

Vonk, W., and L. G. Noordman. 2012. "On the Control of Inferences in Text understanding." In *Comprehension Processes in Reading*, 447–464. Routledge.

Yeung, P. S., C. S. H. Ho, D. W. O. Chan, and K. K. H. Chung. 2016. "A Componential Model of Reading in Chinese." *Learning and Individual Differences* 45: 11–24. https://doi.org/10.1016/j.lindif.2015.11.007.

Zhu, X., C. M. Cheong, G. Y. Li, and J. Wu. 2020. "Primary School Teachers' Conceptions of Reading Comprehension Processes and Its Formulation." *Frontiers in Psychology* 11: 615. https://doi.org/10.3389/fpsyg.2020.00615.

Zwaan, R. A., M. C. Langston, and A. C. Graesser. 1995. "The Construction of Situation Models in Narrative Comprehension: An Event-Indexing Model." *Psychological Science* 6, no. 5: 292–297. https://doi.org/10.1111/j.1467-9280.1995.tb00513.x.