

# Effects of ChatGPT on Korean EFL Learners' Main-Idea Reading Comprehension via Top-Down Processing

Rakhun Kim<sup>†</sup>

Hankuk University of Foreign Studies

---

## ABSTRACT

This study quantitatively and qualitatively investigated the effects of ChatGPT on top-down processing in English reading comprehension among Korean English as a foreign language (EFL) learners. Participants were divided into experimental and control groups of 20 individuals each, completed pretest, immediate posttest, and delayed posttest sessions, and received explicit main idea identification instructions for Finding the Main Idea (FMI) test items from the Korean College Scholastic Aptitude Test. However, only the experimental group had access to ChatGPT. The results showed a statistically significant improvement in top-down processing skills in the experimental group. Additionally, a qualitative analysis of interviews with learners and ChatGPT-human interaction data revealed responses to ChatGPT's use in English reading comprehension, emphasizing its potential benefits and challenges. These findings highlight the significance of balancing technological integration with pedagogical instruction to optimize learning experiences in the Korean EFL context.

**Keywords:** ChatGPT, artificial intelligence, top-down processing, reading comprehension, Korean EFL learners

---

## 1. Introduction

In recent years, a growing body of empirical studies have highlighted the pedagogical potential of artificial intelligence (AI) within English as a Foreign Language (EFL) or English as a Second Language (ESL) context (Bengio et al., 2003; Cho et al., 2014; Lo, 2023; Rahman & Watanobe, 2023; Tlili et al., 2023; Vaswani et al., 2017). Notably, ChatGPT, a generative AI, has emerged as a promising learning tool, functioning as an “educational assistant” to promote the capabilities of human language instructors, with particular emphasis on the potentiality in

---

<sup>†</sup> Corresponding author: kimrhee22@hufs.ac.kr



facilitating the learning of English productive skills (Jeon & Lee, 2023; Kohnke et al., 2023; Shaikh et al., 2023; Shin, 2023b). Nevertheless, the previous research underscores a significant research gap concerning the impacts of generative AI on reading skills, particularly the comprehension of main ideas through the activation of top-down processing by Korean EFL learners (Kim, 2023a).

This research gap demands closer examination, since it is central to our understanding of the broader implications of generative AI on the teaching and learning of higher-order cognitive skill uses for receptive skills within EFL contexts. The present study is motivated by the imperative to address this significant research gap, aiming to contribute to a more comprehensive understanding of the effects of generative AI on the activation of higher-order cognitive abilities for English receptive skills by Korean EFL learners. The following are the research questions:

- (1) Does ChatGPT-assisted instruction affect the ability of Korean EFL learners to engage in top-down processing during English reading comprehension tasks?
- (2) What is the learners' perception in performing English reading comprehension tasks using top-down processing with ChatGPT?
- (3) In what ways do interactions between Korean EFL learners and ChatGPT shape Korean EFL learners' top-down reading comprehension strategies, particularly in the context of identifying main ideas?

## 2. Literature Review

### 2.1. English reading comprehension as a hypothesis confirmation process

It has been generally accepted that reading comprehension is a cognitive procedure to decode semantic information, or proposition, encoded in the text (Kintsch & van Dijk, 1978; Maria & MacGinitie, 1982; Marslen-Wilson & Welsh, 1978; Yang & Jeong, 2008). Since proposition consists of argument structure or subject-predicate relation at the sentence level, learners are required to have a control over linguistic knowledge of vocabulary, syntax relevant with their sentence-grammar semantics (Canale & Swain, 1990). Further, since each sentence is interconnected with different informational significance, proficient readers need to sort out more important proposition from less ones (Kintsch, 1979).

Kintsch (1979) identified the processes involved in bottom-up processing, encompassing word recognition, access to word meanings, understanding of propositional meanings within sentences (Carrell, 1998; Kintsch & Van Dijk, 1978), and syntactic analysis. In contrast, top-down processing incorporates factors such as readers' objectives, world knowledge, and the schema that organizes the text (Dechant, 1991; Goodman, 1982, 1998; Marslen-Wilson & Welsh, 1978; Urquhart & Weir, 1998). The schema is categorized into two types: content schema and formal schema. In this study, the focus is primarily on formal schema, which pertains to the organization of discourse structure (e.g., cause-effect, compare-contrast, problem-solution, etc.) which is often indicated by discourse markers (Halliday & Hasan, 1976; Hildyard & Olson, 1978; Meyer & Freedle, 1984). In this sense, reading comprehension is a parallel process of two directions, namely, top-down and bottom-up (Kintsch, 1979; Rumelhart, 1977). Thus, when foreign language learners engage in text comprehension, they employ both bottom-up and top-down strategies.

As outlined by Urquhart and Weir (1998), top-down involves hypothesis-confirmation processes. This cognitive operation entails readers approaching their reading experiences with specific expectations and ongoing predictions. Should their predictions prove inaccurate during the reading, readers revisit the text, subjecting it to a closer scrutiny to revise their hypotheses (Goodman, 1998). In fact, proficient readers often pay attention to the discourse markers as part of the evidence to confirm or refute their hypothesis (Stauffer, 1969).

Within Korean English as a Foreign Language (EFL) contexts, proficient English learners with high proficiency tend to employ top-down strategies more frequently compared to their lower proficiency counterparts (Bang, 2020; Hong-Nam & Page, 2014). Regrettably, a majority of Korean EFL learners with less proficiency tend to prioritize bottom-up processing (Park, 2018). This may lead learners to over-rely on bottom-up processing, resulting in the failure of reading comprehension (Koda, 2007). Thus, in order to avoid such overreliance on bottom-up processing and to promote balanced reading comprehension processing (i.e., interactive reading; Grabe, 1998), teachers must provide explicit instruction on the activation of top-down processing (Chamot & O'Malley, 1994; Chong & Kim, 2023; Hedge, 2000; Nuttall, 2000; Park, 2018). Furthermore, Korean learners need an interlocutor in the process of hypothesis verification while employing the top-down strategy.

To address these pedagogical requirements, ChatGPT, a generative AI chatbot, has been introduced into educational settings. As will be discussed in the following section, ChatGPT has the technological potential to provide Korean EFL learners

with instructional intervention for bottom-up processing. Moreover, ChatGPT holds the potential to function as both a virtual tutor and an interlocutor for validating learners' hypotheses for top-top processing during reading comprehension. The following subsections will introduce the technological aspects of ChatGPT and its implications in foreign language education.

## 2.2. ChatGPT and English language learning

Technologically, ChatGPT exhibits generative, pre-trained, and transformative characteristics (Goodfellow, Bengio, & Courville, 2016; Krizhevsky, Sutskever, & Hinton, 2017). In its generative aspect, ChatGPT not only comprehends natural human language but also produces highly relevant responses to human counterparts' utterances. Furthermore, it undergoes pretraining (or RLHF; reinforcement learning from human feedback) on extensive, unlabeled, general-domain corpora, employing a large language model (LLM) to facilitate task-oriented dialogue modeling (Bengio et al., 2003; Brown et al., 2020; Vaswani et al., 2017). Lastly, the transformative model represents one of the most advanced deep learning (DL) models in computer science (Brown et al., 2020; Cho et al., 2014; Gers, Schmidhuber, & Cummins, 2000). Combining these three aspects, ChatGPT can be characterized as one of the most efficient conversational AI agents that relies on an LLM and is built upon one of the most well-established DL models. Hence, ChatGPT demonstrates the capability to generate contextually relevant responses to human queries (i.e., prompts), engaging in human-like conversations while exhibiting rudimentary reasoning abilities, particularly in the context of everyday life experiences (Wei et al., 2022; Zhang et al., 2023).

ChatGPT, recognized as one of the most innovative DL models, has drawn significant attention due to its potential applications in educational contexts. As such, a significant body of research has advocated for the pedagogical application of ChatGPT across general educational settings (Lo, 2023; Rahman & Watanobe, 2023; Tlili et al., 2023). These prior studies have consistently arrived at a unanimous consensus, highlighting that ChatGPT enhances student engagement, individualizes learning experiences, and complements the role of educators by functioning as a virtual instructor capable of automatically generating lesson plans, educational resources, and assessment materials.

Moreover, beyond its broader educational applicability, ChatGPT has recently become the focus of growing interest in ESL/EFL research, where it has been

accepted as a promising virtual tutor (Jeon & Lee, 2023). In fact, the comprehensive investigation by Kohnke et al. (2023) unveiled the pedagogical interventions of ChatGPT to be akin to those of human instructors. These interventions include the provision of contextualized word meanings and dictionary definitions, correction of language errors, generation of texts spanning various genres (e.g., email, narratives, or recipes), annotation of texts, paraphrasing of sentences, production of assessment materials, and translation of L2 into L1 expressions. Furthermore, as demonstrated by Shaikh et al. (2023), ChatGPT has demonstrated its effectiveness in providing constructive feedback on learners' spoken and written productions, with a particular emphasis on the evaluation of content, grammatical accuracy, and vocabulary usage. In this context, ChatGPT may provide a venue for dynamic interaction within the learner's zone of proximal development (ZPD) (Palinscar & Brown, 1984).

Nevertheless, the educational implementation of ChatGPT is not without pedagogical limitations. Specifically, as highlighted by Lo (2023), students who engage with ChatGPT within educational contexts have encountered challenges regarding its precision and reliability (e.g., hallucination effect, Alkaiissi & McFarlane, 2023; Bang et al., 2023; Rudolph, Tan, & Tan, 2023). To address such concerns effectively, it is incumbent upon instructors to guide students in validating the information provided by ChatGPT with other authoritative sources, such as reference books or teachers' instruction. In this respect, language instructors should encourage students to critically assess the content generated by ChatGPT (Kasneji et al., 2023; Liu & Ma, 2023; Reinders & Benson, 2017; Szabo, 2023).

### 2.3. ChatGPT and English language learning in the South Korean EFL reading comprehension context

In the South Korean EFL context, initial research on ChatGPT in educational settings predominantly centered around its role in the development of pedagogical materials (Kim, 2023b; Shin, Chung, & Lee, 2023), its application in assessment tasks (Shin, 2023a), and assessments of its technological performance in English reading tests (Kwon & Lee, 2023).

One of the first attempts to explore the instructional effects of ChatGPT on the development of English reading skills among Korean EFL learners was conducted by Kim (2023a). This study delved into the instructional role of ChatGPT, with a particular focus on the development of English receptive skills, especially in the context of reading, by Korean EFL learners. The research specifically explored the

integration of ChatGPT within English reading classrooms, where it functioned as a virtual English tutor, providing real-time assistance to students during their reading exercises. The results of this study revealed a prevalent trend wherein the majority of students primarily sought ChatGPT's support for tasks related to bottom-up processing skills, such as vocabulary, grammar, and L1-L2 translation. In contrast, there was a notable absence of inquiries regarding top-down processing skills, such as the identification of the thesis statement. Consequently, the instructional impact of ChatGPT was predominantly manifested in tasks associated with bottom-up processing, while its influence on tasks necessitating top-down processing remained relatively limited. Given these research findings, it becomes imperative to further explore ChatGPT's potential contribution to the development of higher-order cognitive processing, particularly in main idea comprehension tasks by Korean EFL learners.

### 3. Methods

#### 3.1. Participants

In this study, participants were organized into two distinct groups: The experimental control group. The experimental group consisted of 20 participants ( $n=20$ ), with 12 males and 8 females, averaging 17.6 years of age. They self-reported their English reading proficiency as novice and had, on average, 9.7 years of experience in learning English. The control group also comprised 20 individuals ( $n=20$ ), with 13 males and 7 females, averaging 17.5 years in age. Like the experimental group, they reported their English reading proficiency as novice and had an average of 9.9 years of English learning experience. The participants were enrolled in three different high schools, the one located in Seoul, and the other two in Gyeonggi Province. Students volunteered to participate in this study based upon parental consent, and their participation was financially rewarded, equivalent to 35 U.S. dollars.

The researcher was not involved in the instructional practices, and three English teachers participated in the present study. Their average teaching experience was 13.2 years, and all of them had obtained master's degree of English education. Their participation was financially reimbursed, equivalent to 100 U.S. dollars.

To ensure ethical considerations, parental consent was obtained from the

participants in the control group, just as in the experimental group, and each participant was financially reimbursed for their contribution to the study in Korean currency equivalent to 10 U.S. dollars.

### 3.2. Procedures and instructional treatments

A series of instructional sessions took place between April 4th and June 25th, 2023, encompassing a total of 20 afterschool sessions. Figure 1 outlines a general framework of the present study.

During the instructional sessions, the experimental group, composed of 20 participants, utilized English reading reference book published by Educational Broadcasting System (EBS), a South Korean company, as a part of their regular afterschool English classes, which was specifically targeted for the preparation of the Korean College Scholastic Aptitude Test (K-CSAT).

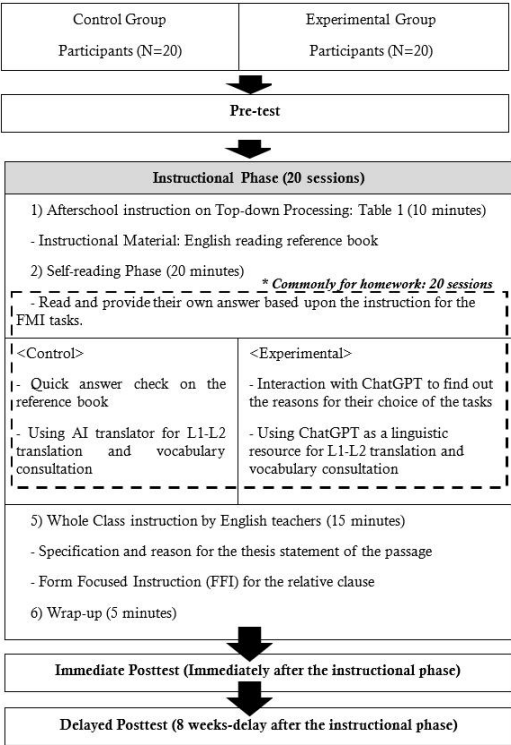


Figure 1. Procedure of the study

The instructional approach predominantly emphasized the activation of top-down processing through the utilization of discourse markers. Thus, as is outlined in Table 1, both experimental and control group participants were exposed to the same instructional treatments that included the implementation of a top-down reading strategy, with a particular emphasis on discourse markers, specifically targeting keywords for identifying the main idea. Initially, a pretest was administered to evaluate the participants' proficiency in English main idea reading comprehension. The control group, also comprising 20 participants, received the same instructional materials and treatments related to top-down reading strategies and discourse markers as their experimental group counterparts.

**Table 1.** Common instructional material across the two groups: List of discourse markers for identifying main idea

Discourse structure	Discourse marker
Thesis Statement & Example	Thesis statement is placed prior to the discourse markers indicating list of examples (e.g., <i>For example, According to ...</i> ).
Deontic Modality Auxiliary	Thesis statement usually incorporates deontic modality auxiliary (e.g., <i>Should, have to...</i> ).
Adversative Transition	Thesis statement usually incorporates adversative transitional words (e.g., <i>However, but ...</i> ).
Causal Transition	Thesis statement usually incorporates causal transitional words (e.g., <i>Consequently, to sum up ...</i> ).

Over the study period, both groups were assigned two tasks as a single homework assignment twice a week, with a total of 40 reading tasks sourced from K-CSAT test sets. These reading tasks involved various question types, including Finding Main Idea (FMI), Finding Factual Information (FFI), Fill-In-The-Blank (FITB), and Identifying Sentence Order (ISO). It is worth noting, as described by Kim (2010), that each of these task types is associated with distinct processing strategies: Specifically, FMI is aligned with top-down processing, FFI corresponds to bottom-up processing, while FITB and ISO tasks involve interactive processing strategies.

What differentiated the experimental group from the control group was whether ChatGPT was available for the afterschool assignment activities. Specifically, participants in the experimental group had guidance in how to use ChatGPT to articulate the reasoning for their answers. They were instructed to upload the reading comprehension text file to the ChatGPT. Learners could do so by taking a picture of the text material or by screen-capturing the relevant text. At this point, questions concerning the bottom-up processing were procrastinated. The prompt to ChatGPT



was delivered in the learners' first language (e.g., Korean), and the scope was restricted to the location of the most informative sentence (i.e., "Show me the most informative sentences."). The term "thesis statement" was not used, because ChatGPT usually provided the first sentence of the paragraph in providing the answer for the relevant question. Learners then asked ChatGPT for the reasoning in choosing the relevant expressions (e.g., "Why is "..." more informative than the other sentence?") and compared ChatGPT's suggestions with their own. If they needed more information for the ChatGPT's reasoning, the interactional process continued with further questions or prompts. Finally, learners were allowed to ask the bottom-up processing related questions on vocabulary or grammatical information.

Conversely, the control group participants did not have access to ChatGPT for afterschool assignments: They were requested to provide their own answers for the questions and the correct answers were then provided.

Following the instructional phase, immediate tests were administered to assess the effects of the instructional intervention. After an interval of eight weeks, delayed posttests were conducted.

### 3.3. Testing instruments

This study employed 20 multiple-choice English reading comprehension questions from K-CSAT during pretest, immediate post-test, and delayed post-test session. Since the primary research focus was to assess learners' top-down processing, five Finding Main Idea (FMI) test items were included as testing items. Additionally, the assessment encompassed 5 Finding Factual Information (FFI) questions to activate bottom-up processing, another set of 5 Finding Main Idea (FMI) questions to stimulate top-down processing, along with 5 Fill-In-The-Blank (FITB) questions and 5 tasks for Identifying Sentence Order (ISO) were included as filler items. All the testing items were selected from nine sets of K-CAST from the years of 2019, 2020, and 2021.

### 3.4. Data collection procedures for the chatting log with ChatGPT and interview

For the qualitative data collection, participants took part in two data collection procedures. First, participants were asked to participate in the interview procedures. This study employed semi-structured interviews. Based upon previous studies on the

use of ChatGPT in language learning contexts (c.f., Kim, 2023a; Seok, 2024), three interview questions were deduced. While performing the interview with the participants, this study employed follow-up question techniques (i.e., probing; Merriam & Tisdell, 2015) to promote learners' free and creative responses on the topic of the use of ChatGPT in performing top-down processing. Table 2 outlines the questions employed in the interview session.

**Table 2.** Interview questions

Interview questions
1. What was a positive aspect of your interaction with ChatGPT in performing the task?
2. What was a negat aspect of your interaction with ChatGPT in performing the task?
3. Do you think ChatGPT was helpful in solving FMI questions? Tell me why.

Second, participants were asked to submit the chatting log with ChatGPT once a week to investigate the interactional pattern between ChatGPT and human learners in discussing how human learners perform top-down processing. As shown in Table 3, based upon the ChatGPT log data, learners' significant statements were deduced and clustered. Two other peer researchers cross-checked the data and adjusted them for certain inaccuracies. Furthermore, to promote the validity of the study, the triangulation method was applied by incorporating the ChatGPT dialogue log, interview data, and three test results (i.e., pre-, immediate post-, delayed post-test).

**Table 3.** Themed clusters and significant statements

Type of processing	Themed cluster	Significant statement
Top-down	Location of thesis statement	<i>"Give me the sentences that are more informative than the others."</i>
	Reasons of the location of thesis statement	<i>"Why is "... more informative than the others?"</i>
	Mismatch between ChatGPT and learners' reasoning for the choice of the thesis statement	<i>"How about... ?"</i>
Bottom-up	Vocabulary	<i>"What is ...?"</i>
	Grammar	<i>"Provide me with the grammatical analysis for ..."</i>
	English-to-Korean Translation	<i>"Provide me with the Korean translation for ..."</i>

### 3.5. Coding and analysis

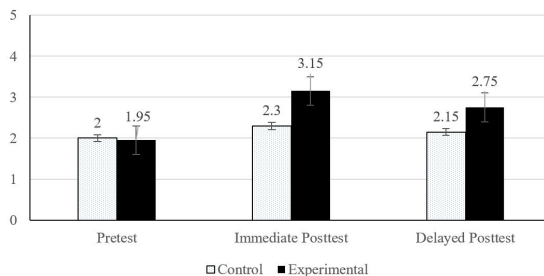
To address the first research question (RQ1), a statistical analysis was conducted on the test results obtained from each testing session. This analysis involved the application of a Two-way Analysis of Variance (ANOVA) utilizing the R software. Notably, the statistical examination excluded the results of filler items, focusing solely on the analysis of Finding Main Idea (FMI) results.

To address the second research question (RQ2), qualitative analysis was employed to assess learners' perspectives. participants were interviewed to evaluate the effectiveness of the top-down processing instruction and to explore the extent to which the utilization of ChatGPT influenced their top-down processing during the test sessions. To account for the third research question (RQ3), participants' interactional log with ChatGPT while performing top-down processing was collected, following each testing session.

## 4. Results and Discussions

### 4.1. Quantitative analysis: Instructional effects of ChatGPT on the improvement of top-down processing in English reading comprehension

To answer RQ1, an analysis of the reading comprehension test on Finding Main Idea (FMI) was carried out to determine whether the experimental and control groups differed in their ability to comprehend English main idea across the three testing sessions. Figure 2 illustrates the mean score differences between the three tests across the two groups.



**Figure 2.** Mean scores of FMI

Table 4 reports the mean score of FMI of the experimental and control group across the three-testing session. It demonstrates that, prior to instruction, all participants in both groups had relatively similar scores in FMI (control: 2.0, experimental: 1.95). A paired sample t-test was conducting on R software to compare the mean score of the two groups (Control Group,  $M=2.0$ ,  $SD= .86$ ; Experimental Group,  $M=1.95$ ,  $SD= .83$ ), and no statistical difference was found between them ( $t(38)= .18$ ,  $p= .82$ ) (R Development Core Team, 2008).

In the two posttest sessions, the experimental group demonstrated a higher mean score than the control group in the two posttest sessions (*immediate posttest*, control: 2.31, experimental: 3.15, *delayed posttest*, control: 2.15, experimental: 2.75). This demonstrates the intervention of ChatGPT played a positive role in the comprehension of main idea.

**Table 4.** Mean score of the K-CSAT reading comprehension tasks for FMI

Group	Pretest (Max = 5)		Immediate posttest (Max = 5)		Delayed posttest (Max = 5)	
	Mean	SD	Mean	SD	Mean	SD
Control	2.0	.86	2.3	.78	2.15	.73
Experimental	1.95	.83	3.15	.92	2.75	.77

To investigate the statistical significance of the instructional effects within and between the groups, a repeated-measure ANOVA was conducted. Statistical significance was found in the test session ( $F(2, 117) = 12.13$ ,  $p < .001$ ,  $\eta^2 = .14$ ) and in the treatment groups ( $F(1, 38) = 14.47$ ,  $p < .001$ ,  $\eta^2 = .08$ ). Also, there was a significance for the interaction effect of the groups with the test sessions ( $F(2, 117) = 5.54$ ,  $p = .005$ ,  $\eta^2 = .07$ ).

The mean differences were found in the three test sessions across the instructional group-namely, the experimental group ( $F(2, 19) = 15.67$ ,  $p < .001$ ). No statistical difference was found in the control group ( $F(2, 19) = .69$ ,  $p = .502$ ). The result of the between-group analysis of the pretest confirmed that the two groups were homogenous prior to instruction ( $F(1, 38) = .72$ ,  $p = .82$ ). The effect size, as measured by Cohen’s d was .9, indicating a large effect. However, this result showed a statistical difference between the two groups in the immediate posttest ( $F(1, 38) = 1.97$ ,  $p < .001$ ). However, the result of the between-group analysis did not confirm a statistical difference in the delayed posttest ( $F(1, 38) = 1.37$ ,  $p = .21$ ). This

demonstrated that the immediate instructional effect was evident, but the delayed instructional effect was not found.

#### 4.2. Qualitative analysis: Learners' perception in performing English reading comprehension tasks using top-down processing with ChatGPT

To answer RQ2, the participants were asked how they felt about using ChatGPT after learners finished the instructional sessions. The followings are the excerpts from the interview entry. Each of the interview data were translated from Korean to English by the researcher.

In general, participants described ChatGPT as an amiable, user-friendly, and naturally intuitive mode of communication.

*"I believe that having a chat with ChatGPT was fun. Actually, it sometimes gave me some warm encouragements. I liked that!"* (Participant H, Interview entry log number 25)

They also mentioned that ChatGPT was useful for difficult vocabulary and L1-L2 translations, which surpassed the efficacy of electronic dictionaries or machine translation tools. This aligns with the findings of Kim (2023a).

*"When I used a (company name) translator, it usually gave me a lot of weird translation. It was far from usable, since the translation was out of context. However, the translation that ChatGPT provided was ... more natural."* (Participant J, Interview entry log number 31)

In contrast, the participants noted negative experiences. For example, they reported that the responses generated by ChatGPT were excessively lengthy, posing significant challenges in comprehension. Furthermore, the tone adopted by ChatGPT was critiqued for being perceived as overly assertive, bearing a resemblance to that of an academic expert rather than an affable conversational partner.

*"ChatGPT was boring. He always gave me a lengthy words. He talked like a professor, and it sounded quite arrogant."* (Participant B, Interview entry log number 4)

Finally, learners conveyed that they enjoyed the interaction with the Chat GPT at the outset, but the interactions became tiresome due to primacy/recency effect (Sousa, 2005).

*“At first, using ChatGPT was quite fun and exciting. You know, ChatGPT is a symbol of AI technology. But as I continue to use it, I got used to it, and the excitement that I felt in the beginning of the lesson finally diminished.”* (Participant F, Interview entry log number 19)

These findings underscore the need for a more detailed approach to the incorporation of ChatGPT in English as a Foreign Language learning (EFL) contexts to promote instructional effectiveness. While the participants acknowledged that ChatGPT had potential to effectively contribute to the enhancement of reading comprehension skills, they agreed that ChatGPT should be more carefully adjusted for the pedagogical purposes.

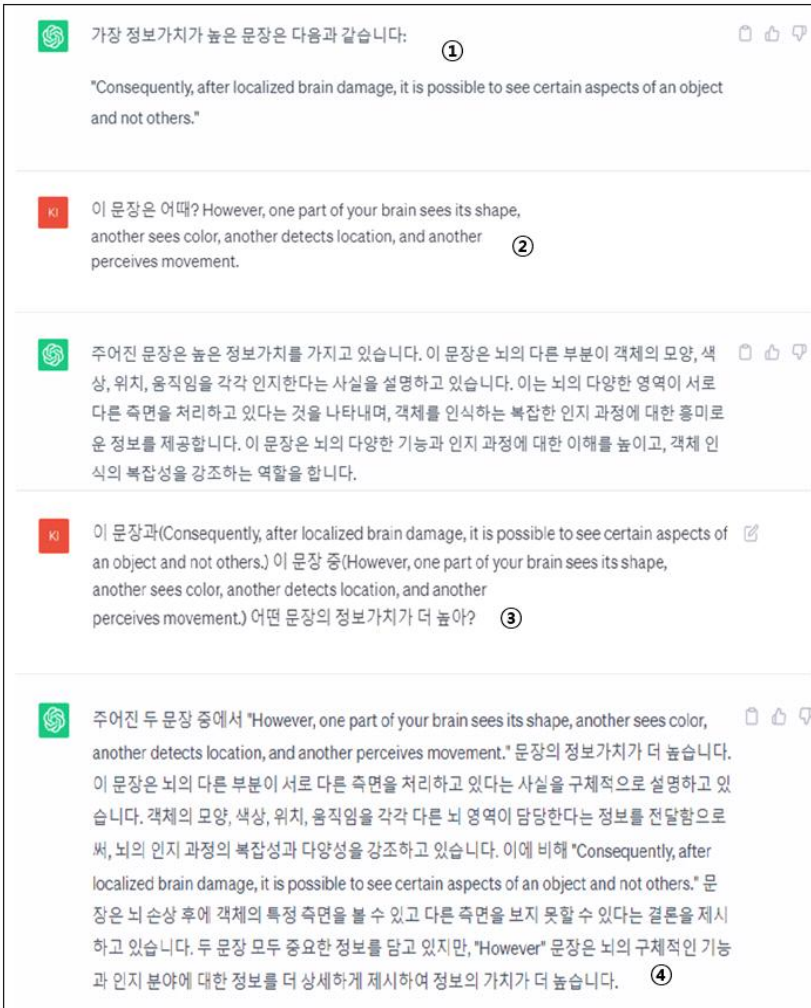
To ensure ChatGPT’s effectiveness in an instructional role as a virtual tutor, the study emphasized the importance of providing comprehensive pedagogical instructions at the prompt-level, underscoring the need for a balanced integration of technological tools and pedagogical strategies to optimize the learning experience. To do so, this study focused on how ChatGPT and human participants interacted when there was a discrepancy between ChatGPT’s pedagogical suggestions and human learners’ reasoning.

#### 4.3. ChatGPT log data analysis: Discrepancy between ChatGPT’s suggestions and learners’ reasoning

To account for RQ3, the interactional log data between ChatGPT and the human participants in the experimental group were analyzed. Based upon the interactional log between ChatGPT and the human participants, the interactions mainly focused on dealing with the discrepancies between the machine’s suggestions and learners’ reasoning for their choices.

First, several instances were identified in which ChatGPT’s suggestions diverged from the learners’ supporting evidence for the main idea: Notably, when such discrepancies emerged, ChatGPT occasionally adjusted its initial suggestion to align with the learners’ viewpoint. This phenomenon is exemplified in Figure 3, which illustrates the evolution of a discussion between ChatGPT and a human participant. Initially, when the learner inquired about the most informative sentence, ChatGPT provided a suggestion, as shown in (①) (e.g., *Consequently, after localized brain damage, it is possible to see certain aspects of an object and not others.*). However, as depicted in (②), the learner challenged ChatGPT’s suggestion by offering alternative

candidates (e.g., *However, one part of your brain sees its shape, another sees color, another detects location, and another perceives movement.*). Subsequently, with ChatGPT acknowledging the validity of the learner’s suggestion, the learner proceeded to inquire about a comparison between the two viewpoints, as seen in (3). Ultimately, as demonstrated in (4), ChatGPT modified its initial suggestion to align with the learner’s viewpoint. This clearly indicates a certain degree of adaptability in response to learners’ reasoning.

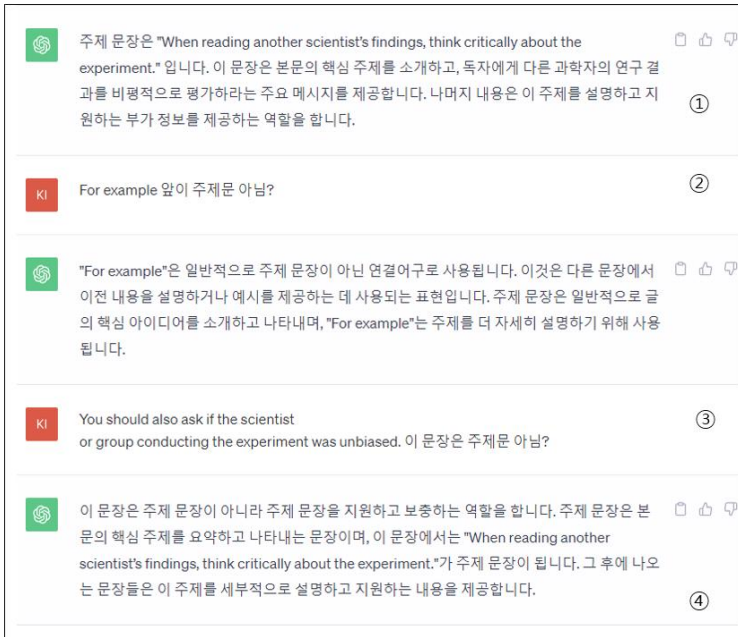


**Figure 3.** Example of mismatch between ChatGPT’s initial suggestion and learner’s response (1): Final alignment with learner’s suggestion

Secondly, although the human learners provided the correct thesis statement, ChatGPT occasionally did not change its initial suggestion which was accurate in identifying thesis statement. This observation is illustrated in Figure 4, where ChatGPT initially provided a sentence from the passage as a thesis statement (e.g., “*When reading another scientist’s findings, think critically about the experiment.*”) (①). Nevertheless, the learner, guided by the instruction from the human teacher, challenged ChatGPT’s recommendation: As ② and ③ demonstrates, the participant had learned that a thesis statement is typically positioned before “For example” (e.g., “*Being unbiased means that you have no special interest in the outcome of the experiment. For example...*”) and exhibits deontic modality (e.g., “*You should also ask if the scientist or group conducting the experiment was unbiased.*”). Despite this, ChatGPT rejected the learner’s proposal, asserting that the suggested thesis statement of the human learner was merely a supporting sentence (④).

It is noteworthy that the persistent incongruity between ChatGPT and human learners led to a robust and dynamic discussion among the learner, their peers, and the instructor. Learners exhibited a clear preference for the instructor’s guidance in the realm of main idea identification, giving it precedence over ChatGPT’s suggestions. This preference was accompanied by a notable reluctance to embrace ChatGPT’s recommendations. In fact, the disparity between the pedagogical interventions provided by the instructor and those offered by ChatGPT has acted as a catalytic force, promoting internal cognitive processing within the learner. While it is true that this mismatch at times engendered confusion in the learner’s understanding, the instructor’s instructions consistently prevailed over ChatGPT’s suggestions. Consequently, this served as the foundation for the learner’s active and voluntary participation in subsequent discussions with both fellow students and the instructor.





**Figure 4.** Example of mismatch between ChatGPT’s initial suggestion and learner’s response (2): Persistent mismatch with learner’s suggestion

The constant disparity between ChatGPT and human learners, particularly in the context of thesis statement identification, poses two important implications. Firstly, the incorporation of ChatGPT for enhancing reading comprehension offers a promising avenue for learners to actively engage in a discussion to critically assess ChatGPT’s opinions. When viewed through the lenses of social interaction theory, this scenario can be regarded as an instance in which learners, drawing on their pre-existing knowledge of pedagogical strategies, take an active role in the reciprocal teaching and learning processes (Palinscar & Brown, 1984).

Second, the interaction data strongly suggests that ChatGPT should be perceived not as an authoritative instructor but rather as a conversational interlocutor which is prone to making errors (Jeon & Lee, 2023). While ChatGPT can assume a few pedagogical roles in foreign language learning contexts, it is crucial to retain the human teacher as the paramount authority. This becomes especially vital when considering human teacher’s profound pedagogical expertise: Although ChatGPT may be proficient at delivering credible responses to nearly all learner queries, the likelihood of ensuring objectivity, as exemplified by the “hallucination effect,” is significantly lower in comparison to human instructors (Kim, 2023b).

## 5. Conclusion

### 5.1. Pedagogical implications

This study addresses three key research questions regarding the impact of ChatGPT-assisted instruction on top-down processing in English reading comprehension by Korean EFL learners.

It found that instructional intervention from ChatGPT had a significant effect on the participants' ability to identify the main idea within FMI test items (RQ1). Moreover, in qualitative analyses (RQ2 & RQ3), human participants engaged in dialogue with ChatGPT, focusing on logical evidence related to the identification of the main idea in reading comprehension tasks. The interactive attempt between ChatGPT and human learners promoted an active process of top-down processing, as learners sought to verify the congruence of ChatGPT's suggestions with their own expectations. Upon finding such alignment, the discussion shifted towards bottom-up processing. Otherwise, in cases where disparities arose between their chosen evidence and ChatGPT's suggestions, learners undertook efforts to substantiate their decision. This dialogic exchange with ChatGPT emerged as a pivotal factor in developing top-down processing skills (Palinscar & Brown, 1984).

It should be highlighted that the application of ChatGPT, one of the sub-kinds of artificial intelligence technology, does not automatically lead to the enhancement of pedagogical efficiency. Instead, the opportunity to articulate the reasoning for the students' choice of answer seems to be one of the key components for the development of their reading comprehension skills. However, human teachers are practically not able to provide individual language learners with optimized feedback. Thus, the pedagogical value of ChatGPT lies not in whether it is applied to the teaching and learning contexts of language learning, but in how it is applied to the teaching and learning contexts of English language learning: To maximize the pedagogical potentiality of ChatGPT, thus, it is required for the language instructors to contextualize the use of ChatGPT in the teaching and learning processes.

In summary, this study provides quantitative evidence that ChatGPT-assisted instruction positively impacts the identification of main ideas in reading comprehension and demonstrates how the utilization of ChatGPT leads to the activation of top-down processing among Korean EFL learners in a qualitative

manner. This study is important for being the first to experimentally examine the instructional intervention of ChatGPT on the activation of top-down processing among Korean EFL learners, especially for receptive skill (e.g., reading comprehension). The research findings reported in the present study shed light on the pedagogical potential of ChatGPT for English language learning in both EFL and ESL contexts.

## 5.2. Limitations

The present research is not without limitations. First, the research encountered practical challenges in achieving a high level of test-retest reliability. Despite our concerted efforts to ensure test-retest reliability, we could only achieve a marginal consistency level. Thus, it is essential to fine-tune the testing materials to enhance the reliability of the assessments across the testing sessions. Second, our study incorporated a stepwise pedagogical intervention process with ChatGPT, but some participants deviated from the intended process, focusing more on bottom-up processing, such as asking for vocabulary and L1-L2 translation. They did not fully apply the top-down processing techniques as instructed. In subsequent research, greater attention should be given to the meticulous planning of the language learners' engagement with ChatGPT to ensure they adhere to the instructor's pedagogical guidance. Furthermore, the small sample size of the present study significantly limits its generalizability. Finally, the present study could not find a delayed instructional effect of ChatGPT on reading comprehension.

Future research is necessary to contribute to a more comprehensive understanding of the practical application of ChatGPT-assisted language learning within Korean EFL contexts. Furthermore, it is imperative to explore pedagogical interventions involving ChatGPT for interactive reading processes (e.g., FITB), which can provide a deeper understanding of how ChatGPT can effectively promote learners' reading comprehension proficiency for the activation of both bottom-up and top-down processing to guarantee long-term instructional effects over the delayed posttest session.

## References

- Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus*, *15*(2).

- Bang, J. (2020). The comparison of metacognitive awareness of reading strategy use among gifted learners in science, mathematics, and general middle school learners during english reading. *Studies in English Language & Literature*, 46(1), 125-145.
- Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., Lovenia, H., Ji, Z., Yu, T., Chung, W., V. Do, Q., Xu, Y., & Fung, P. (2023). A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. *arXiv preprint arXiv:2302.04023*.
- Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, 3, 1137-1155.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.
- Canale, M., & Swain, M. (1990). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Carrell, P. L. (1998). Can reading strategies be successfully taught?. *Australian Review of Applied Linguistics*, 21(1), 1-20.
- Chamot, A. U., & O'malley, J. M. (1994). *The CALLA handbook: Implementing the cognitive academic language learning approach*. New York: Longman.
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.
- Chong, I., & Kim, H. (2023). Effects of explicit reading strategy instruction on high and low English proficiency learners. *Journal of Learner-Centered Curriculum and Instruction*, 23(4), 715-731.
- Dechant, E. (1991). *Understanding and teaching reading: An interactive model*. New Jersey: Lawrence Erlbaum.
- Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12(10), 2451-2471.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge: MIT Press.
- Goodman, K. (1982). *Language and literacy*. Boston: Routledge & Kegan Paul.
- Goodman, K. (1998). The reading process. In P. Carrell, J. Devine, & D. Eskey

- (Eds), *Interactive approaches to second language reading* (pp. 11-21). Cambridge: Cambridge University Press.
- Grabe, W. (1998). Reassessing the term interactive. In P. Carrell, J. Devine & D. Eskey, (Eds.), *Interactive approaches to second language reading* (pp. 56-70). Cambridge: Cambridge University Press.
- Halliday, M. A., & Hasan. R. (1976). *Cohesion in English*. London: Longman.
- Hildyard, A., & Olson, D. R. (1978). Memory and inference in the comprehension of oral and written discourse. *Discourse Processes*, 1(2), 91-117.
- Hedge, T. (2000). *Teaching and learning in the language classroom*. Oxford: Oxford University Press
- Hong-Nam, K., & Page, L. (2014). Investigating metacognitive awareness and reading strategy use of EFL Korean university students. *Reading Psychology*, 35(3), 195-220.
- Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, 1-20.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.
- Kim, R. (2023a). Effects of ChatGPT on the cognitive processing of K-CSAT English reading tasks by Korean high school learners: A preliminary study. *Secondary English Education*, 16(2), 179-205.
- Kim, T. (2023b). A field study on the development of teaching materials for secondary English and its use of ChatGPT. *Secondary English Education*, 16(2), 207-218.
- Kintsch, W. (1979). On modeling comprehension. *Educational Psychologist*, 14(1), 3-14.
- Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363-394.
- Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning*, 57(1), 1-44.
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537-550.

- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.
- Kwon, S. K., & Lee, Y. (2023). Investigating the performance of generative AI ChatGPT's reading comprehension ability. *Journal of the Korea English Education Society*, 22(2), 147-172.
- Liu, G., & Ma, C. (2023). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching*, 1-14.
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410.
- Maria, K., & MacGinitie, W. H. (1982). Reading comprehension disabilities: Knowledge structures and non-accommodating text processing strategies. *Annals of Dyslexia*, 32(1), 31-59.
- Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10(1), 29-63.
- Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.
- Meyer, B. J. F., & Freedle, R. O. (1984). Effects of discourse type on recall. *American Educational Research Journal*, 21, 121-143.
- Nuttall, C. (2000). *Teaching reading skills in a foreign language*. London: Heinemann.
- Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1(2), 117-175.
- Park, J. (2018). The impact of EFL students' English proficiency on reading strategies and usage of annotation tools. *English Language Assessment*, 13, 85-120.
- R Development Core Team. (2008). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.Rproject.org>.
- Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences*, 13(9), 5783.
- Reinders, H., & Benson, P. (2017). Research agenda: Language learning beyond the classroom. *Language Teaching*, 50(4), 561-578.
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching*, 6(1), 1-22.
- Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic

- (Ed.). *Attention and performance* (Vol. 6). Hillsdale: Erlbaum.
- Seok, W. (2024). *Korean secondary English teachers' use of ChatGPT as teaching and learning tool: Practices and perceptions* [Unpublished MS dissertation]. Korea National University of Education.
- Shaikh, S., Yayilgan, S. Y., Klimova, B., & Pikhart, M. (2023). Assessing the usability of ChatGPT for formal English language learning. *European Journal of Investigation in Health, Psychology and Education*, 13(9), 1937-1960.
- Shin, D. (2023a). A case study on english test item development training for secondary school teachers using AI tools: Focusing on ChatGPT. *Language Research*, 59(1), 21-42.
- Shin, D. (2023b). Utilizing ChatGPT in guided writing activities. *Journal of the Korea English Education Society*, 22(2), 197-217.
- Shin, D., Jung, H., & Lee, Y. (2023). Exploring the potential of using ChatGPT as a content-based English learning and teaching tool. *Journal of the Korea English Education Society*, 22(1), 171-192.
- Sousa, D. A. (2005). *How the brain learns* (3rd ed.). Thousand Oaks, CA: Corwin Press/Sage Publications.
- Stauffer, R. G. (1969). *Directing reading maturity as a cognitive process*. New York: Harper & Row.
- Szabo, A. (2023). ChatGPT is a breakthrough in science and education but fails a test in sports and exercise psychology. *Baltic Journal of Sport and Health Sciences*, 1(128), 25-40.
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1), 15.
- Urquhart, A. H. & Weir, C. (1998). *Reading in a second language: Process, product and practice*. London: Longman.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, N., Jones, L., Gomez, N., A., Kaiser, L., & Polosukhin, I., (2017). Attention is all you need. In *Advances in neural information processing systems*, 5998-6008.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E., Le, Q., & Zhou, D. (2022). Chain of thought prompting elicits reasoning in large language models. *arXiv preprint arXiv:2201.11903*.
- Yang, H., & Jeong, Y. (2008). *Understanding educational grammar*. Seoul: Hankook Publishing.
- Zhang, C., Zhang, C., Li, C., Qiao, Y., Zheng, S., Kumar, S. D., Zhang, M., Uk

Kim, J., Kim, S. T., Choi, J., Park, G. M., Bae, S. H., Lee, L. H., Hui, P., Kweon, I. S., & Hong, C. (2023). One small step for generative AI, one giant leap for AGI: A complete survey on ChatGPT in AIGC era. *arXiv preprint arXiv:2304.0648*.

Rakhun Kim

Adjunct Professor

Department of English Education

Hankuk University of Foreign Studies

107, Imun-ro, Dongdaemun-gu, Seoul 02450, Korea

E-mail: kimrhee22@hufs.ac.kr

Received: November 4, 2023

Revised version received: April 15, 2024

Accepted: April 24, 2024