An investigation of novice ESL writers' cognitive processes and strategy use of paraphrasing

Emily Di Zhang

Correspondence: yb97712@um.edu.mo
University of Macau, Macau, China

Abstract

Source use competence is becoming increasingly important in English-medium universities, and inability to appropriately use reading sources leads to plagiarism which has serious punishable consequences. As one effective strategy to help academic writers to avoid plagiarism, paraphrasing is highly recommended for students to master. However, studies on paraphrasing are product-oriented, and few examined the processes and strategies of paraphrasing. The elusive construct of paraphrasing exerted a negative influence on the instruction and assessment of paraphrasing. Only specific paraphrasing strategies were examined in prior studies, and little has been done on more general strategies like cognitive and metacognitive strategies. Moreover, as for the studies of learner strategy and language performance, no consensus has been reached about the relationship between strategy use and language performance. Prior studies mainly adopted Purpura's questionnaire (1997) with cognitive and metacognitive strategies excluding other important learner strategies, and prior studies mainly examined language performance in non-integrated tasks. Questionnaire and interview were employed in previous studies, but few has adopted think-aloud protocols to capture test takers' online cognitive processes. To address the research gap, the present study purports to shed light on the cognitive processes and strategy use in the paraphrasing task with the method of think-aloud protocols and strategy use questionnaire. A total of 212 first-year non-English-major college students were recruited to respond to a seven-item paraphrasing task. Think-aloud protocol was conducted to capture test takers' thinking processes while responding to the task. In addition, test takers' strategy use in paraphrasing was elicited by a paraphrasing strategy use inventory. Findings of the study are as follows: test takers employed numerous strategies including cognitive, metacognitive, compensation, and affective strategies with a variety of sub-strategies; metacognitive strategies are significantly correlated with cognitive strategies; more proficient test takers used a smaller number of strategies and more metacognitive strategies. However, structural equation modeling analysis reported that all those strategies exert a negligible effect on participants' paraphrasing performance. Other factors which might contribute to the language performance were analyzed, including language knowledge, test methods, personal attributes, and errors of measurement.

Keywords: Paraphrasing, Cognitive process, Strategy use, Paraphrasing performance

Introduction

A growing body of literature addressing academic literacy activities shows that source use competence is frequently expected by the English-medium universities (Bridgeman & Carlson, 1983; Hale et al., 1996), which also proves to be one of the most demanding academic abilities for L2 writers (Hirvela & Du, 2013; Kucer, 1985; McGinley, 1992; Spivey, 1990). Inability to appropriately use reading sources leads to plagiarism which has serious punishable consequences (Pennycook, 1996; Weigle, 2004; Shi, 2010; Gebril, 2010; Hirvela & Du, 2013). To help students stay away from plagiarism, paraphrasing is one of the most highly recommended skill for students to master (Campbell, 1990; Currie, 1998; Howard, 1996; Hyland, 2001; Johns & Mayes, 1990; Keck, 2006).

Paraphrasing is defined as faithful recasting of the original texts with credits to the original author (Mori, 2019; Shi & Dong, 2018). Researchers took a product-oriented approach to building the taxonomies of paraphrasing types based on the ratio of verbatim source use (Shi, 2016; Keck, 2006) and linguistic changes (Burstein et al., 2012). Despite that product-oriented studies on paraphrasing offered insights into the linguistic features of paraphrasing products, little information is provided about how students paraphrase. Geared towards the processes and strategy use in paraphrasing, a handful of process-oriented studies were conducted, and found that paraphrasing skill experiences a developmental path: novice ESL writers used paraphrasing mainly as a knowledge telling strategy (Hirvela & Du, 2013), while advanced ESL writers used paraphrasing as a knowledge transforming strategy (Shi & Dong, 2018). Paraphrasing is found to be influenced by L1 culture and writing experience (Shi & Dong, 2018). However, the three process-oriented studies on paraphrasing revolved around the functions of paraphrasing either as a knowledge telling strategy or as a knowledge transforming strategy, and failed to make focal participants' online cognitive processes and conscious strategy use in paraphrasing as well as the effect of strategy use on the paraphrasing performance.

Though no studies have been conducted to investigate the relationship between paraphrasing strategy use and paraphrasing performance, the effect of strategy use on language performance has been a hotly discussed topic and there is still no consensus in the literature. A proliferation of studies was undertaken on the relationship between learner strategies and learners' performance (Purpura, 1997; Song & Cheng, 2008; Phakiti, 2003; Zhang & Zhang, 2013; Zhang, Goh, & Kunnan, 2014; Cai & Kunnan, 2020; Yang & Plakens, 2012). Purpura (1997) developed a strategy use questionnaire based on the theory of human information processing, and the structural equation modeling (SEM) analysis showed that cognitive strategies exerted no direct and significant effect on test takers' test performance. Phakiti (2003) drew on Purpura's questionnaire and adapted it to the reading strategies. The MANOVA suggested that both cognitive and metacognitive strategies had weak and significant effect on reading performance. Given the mixed findings on this topic, Song and Cheng (2008) adopted Purpura's questionnaire and ran multiple regression analyses for the relationship of test takers' strategy use and their performance on a proficiency test. They found that only some cognitive strategies had significant impact on test takers' performance and the effect was also small. To know which areas of the test performance were significantly influenced by cognitive or metacognitive strategies, Zhang and Zhang (2013) conducted SEM analysis and found that cognitive strategies exerted no significant effect on test takers' reading performance, consistent with Purpura's finding (Purpura, 1997). Given that the

measurement instrument was Purpura's questionnaire or its adapted version, which excluded other important strategies like compensation and affective strategies, Yang and Plakens (2012) developed a strategy use inventory for the integrated writing task. SEM analysis suggested that discourse synthesis strategies exerted a significant effect on test takers' writing performance. To see if the finding of the relationship between strategy use and language performance can generalize across samples, Zhang, Goh, and Kunnan (2014) conducted multi-sample SEM analysis, and they found that cognitive and meta-cognitive strategies collectively had a significant effect on test takers' reading performance. However, the specific effects of cognitive and metacognitive strategies on the reading performance were not known respectively. Cai and Kunnan (2020) shed light on the mechanism of how strategy use affects language performance, and the multi-layered moderation analysis showed that the effect of strategy use ability on nursing English reading performance fluctuated in a down-up-down pattern mediated by students' language knowledge.

Considering the research gaps in the studies of paraphrasing strategies and the studies of the relationship between strategy use and language performance, the present study purports to uncover test takers' cognitive processes and paraphrasing strategies through think-aloud protocols and paraphrasing strategy use inventory, and then examine the relationship between paraphrasing strategies and paraphrasing performance. This study is of great significance. Firstly, it adds to the understanding of paraphrasing processes and strategies from test takers' perspective, which can generate more valid interpretation of paraphrasing strategies. Secondly, studying the processes and strategy use is one important channel of shedding light on the underlying construct of the task (Bachman, 2002), so that important implications can be generated to guide paraphrasing instruction and assessment. Thirdly, the investigation of the relationship between novice ESL writers' strategy use and paraphrasing performance can advance the understanding of the relationship between learner strategy and language performance, which is a controversial topic with mixed findings.

Literature review

Processes and strategies of paraphrasing

Some researchers explored paraphrasing from a product-oriented perspective and made attempts to build the taxonomies of paraphrasing types (Shi, 2016; Keck, 2006; Burstein et al., 2012). Shi (2016) examined the effect of first language and writing task on Chinese and native English undergraduate students' textual borrowing practices. In the coding scheme of textual borrowing, she designed a taxonomy of paraphrasing based on the extent of modification of the source texts, including "copied," "slightly modified by adding or deleting words or using synonyms for content words," and "closely paraphrased by reformulating syntax or changing the wording of the original text." However, the relativistic scale makes it hard to distinguish different categories. Also, there is no "total paraphrasing" which specifies the criteria of successful paraphrasing. Hence, it is pivotal to establish a more reliable and comprehensive taxonomy of paraphrasing. Drawing on Shi's taxonomy (2016), Keck (2006) investigated the use of paraphrasing in L1 and L2 writers' summary writing. She firstly put forward the construct of "attempted paraphrases", excluding those superficial linguistic changes from the taxonomy of

paraphrase types like punctuation. She also specified unique and general links, with the former being related to verbatim source use. The fundamental differences among the paraphrase categories lie in the use of unique links. Four types of paraphrasing were specified, which are "near copy," "minimal revision," "moderate revision," and "substantial revision." The finding for her study was reported that L1 writers were engaged with more "moderate revision" and "substantial revision" and fewer "near copy" than L2 writers. Both Shi (2016) and Keck (2006) viewed paraphrasing as a strategy of textual borrowing in writing tasks, and the strategies are too general to capture a full understanding of the construct of paraphrasing in and of itself. Gearing toward the specific linguistic changes in paraphrasing, Burstein et al. (2012) examined the native and non-native test takers' paraphrasing strategies in TOEFL integrated writing task, which assesses test takers' ability to paraphrase the contrasting viewpoints in reading and listening stimuli. They designed an annotation to gain the understanding of native and non-native test takers' strategies of paraphrasing and how these strategies affect their scores in the writing task. The annotation scheme is based on linguistic changes of the original texts, including "lexical paraphrasing", "syntactic paraphrasing", "conceptual paraphrasing", and "global paraphrasing" with a number of fine-grained sub-categories. Though this taxonomy provides a comprehensive picture of paraphrasing and adds to our understanding to paraphrasing, it cannot provide information about the mental processes and strategy use of paraphrasing from the perspective of paraphrase writers.

To uncover the processes of paraphrasing, Hirvela and Du (2013) conducted think-aloud protocols and text-based interview to capture two Chinese undergraduate students' understanding of the purposes and functions of paraphrasing and how such understanding affects their paraphrasing practices. It is reported that the two participants saw paraphrasing as a knowledge telling strategy, so they excelled at rephrasing texts in decontextualized paraphrasing exercises, while they assigned little rhetorical value to paraphrasing as a knowledge transforming strategy, thus struggling to paraphrase source texts to develop their own arguments in research paper writing. Considering that novice ESL writers' limited language proficiency and experience in academic writing, Shi, Fazel, and Kowkabi (2018) directed attention to the advanced graduate students' paraphrasing practices. Textual analyses and text-based interview were performed to analyze how the participants paraphrased in their academic papers. They found that advanced graduate students used paraphrasing to incorporate source texts into their writing by syntactic restructuring, interpretation, selective representation of relevant source texts, and addition of other source texts from prior reading, suggesting that advanced graduate students were engaged with transforming knowledge in their paraphrases. The findings from Hirvela and Du (2013) and Shi, Fazel, and Kowkabi (2018) provided empirical evidence for what Currie noted that the paraphrasing skill undergoes a developmental path: students gradually make the transition from the knowledge telling stage to the knowledge transforming stage. To examine the difference between L1 and L2 paraphrases, Shi and Dong (2018) went further to perform textual analyses and text-based interview to compare Chinese participants' Chinese and English paraphrases. The finding is that Chinese paraphrases contained significantly more textual borrowing and fewer acknowledgements than English paraphrases. Content recontextualization strategies were found in both L1 and L2 paraphrases, in which English paraphrases featured in selecting information, while Chinese paraphrases were characterized in using one's

own interpretation and adding or extending ideas. The study showed that L1 culture and writing conventions have a role to play in paraphrasing practices. However, all the three process-oriented studies on paraphrasing revolved around the functions of paraphrasing either as a knowledge telling strategy or as a knowledge transforming strategy, and failed to make focal participants' online cognitive processes and conscious strategy use in making full use of their available resources to achieve their goals in paraphrasing.

Research gaps are detected in the literature. Firstly, while studies on paraphrasing either revolved around the textual features of paraphrasing products or the functions of paraphrasing, little is known about the cognitive processes and strategy use of paraphrasing from students' perspectives and the relationship between paraphrasing strategies and paraphrasing performance. Secondly, the construct of paraphrasing is elusive, and the criteria of successful paraphrasing differ among studies. Thus, it is difficult to teach and assess this important ability. Thirdly, the process-oriented studies on paraphrasing strategies are data-driven and strategies are specific to the task. Scant attention is paid to the general strategies which can transfer across tasks. To address the research gaps, the present study seeks to adopt think-aloud protocols and questionnaire of strategy use inventory to elicit novice ESL writers' online cognitive processes and strategy use of paraphrasing. SEM is performed to analyze the relationship between paraphrasing strategy use and paraphrasing performance.

Learner strategy and language performance

As was argued by Cohen (1998), language learner strategies are generally categorized into two types: language learning and language use strategies. The former is adopted to promote learning, while the latter is employed to optimize language performance in some specific contexts. There has been a proliferation of studies on the relationship between learner strategies and learners' language performance (Purpura, 1997; Song & Cheng, 2008; Phakiti, 2003; Zhang & Zhang, 2013; Zhang, Goh, & Kunnan , 2014; Cai & Kunnan , 2020; Yang & Plakens, 2012). The methods are primarily questionnaires and some adopted mix-methods approach with questionnaire and interview. Language performance was assessed in many different ways, such as placement examinations, achievement tests, proficiency tests, and self-rating of language proficiency. The analysis methods mainly include correlation, multiple regression, and SEM.

Grounded in human information-processing theory, Purpura (1997) developed a questionnaire for cognitive and metacognitive strategies and examined the relationship between 1382 test takers' language use strategies and their performance on the FCE Anchor Test, a proficiency test developed by UCLES, which assesses ESL learners' language ability. The cognitive strategies include three variables: comprehending, memory, and retrieval strategies, while metacognitive strategies have one variable, that is, assessment. The construct of the language proficiency test includes reading ability and grammar ability. SEM analyses showed that both metacognitive and cognitive strategy use exerted no direct effect on language performance. Memory strategies had a significant and negative effect on grammar ability, and retrieval strategies had a significant and positive effect on grammar ability. Metacognitive strategies had a direct and significantly positive relationship to cognitive strategies. Drawing on Purpura's questionnaires of cognitive and metacognitive strategies with modification to suit reading test, Phakiti

(2003) used MANOVA to examine the relationship between 384 Thai test takers' strategy use and their reading test performance in a final examination of an English course, and the reading ability is measured by gap-filling cloze and reading comprehension items. The cognitive strategies include comprehending and retrieval strategies, while the metacognitive strategies include planning and monitoring strategies. The finding was reported that cognitive and metacognitive strategies both had a positive but weak relationship with reading performance. He also found that more proficient students used more metacognitive strategies, which had a high correlation with cognitive strategies, corroborating Purpura's finding (1997) that metacognitive strategies had an executive impact on cognitive strategies. Given that there is no consensus on the relationship between strategy use and language performance, Song and Cheng (2008) performed multiple regression to analyze the relationship between 121 ESL learners' strategy use elicited by Purpura's questionnaire (1997) and their language performance on CET-4 measureing general English proficiency. It was reported that participants used more metacognitive strategies than cognitive strategies. Memory and retrieval strategies as subscales of cognitive strategies significantly influenced the language performance on CET-4, though the effect was small. This study shows that strategy use can only contribute to a small portion of the variances of CET-4, and not all strategies related to the test performance. They also found that though metacognitive strategies were more frequently used by test takers, they had no effect on their test performance. Song and Cheng's study (2008) corroborated with Purpura's finding (1997) that metacognitive strategies had no significant effect on language performance, and confirmed Phakiti's finding (Phakiti, 2003) that cognitive strategies had a small but significant effect on language performance. It contributed to the literature by finding that only some cognitive strategies had significant impact on test takers' language performance, which showcased the complexity of the relationship between strategy use and language performance. However, it did not specify what areas of the test were influenced by the memory and retrieval strategies. And moreover, it failed to provide a nuanced understanding of the effect of metacognitive strategies on the test performance. To fill in the gap, Zhang and Zhang (2013) studied the relationship between 209 Chinese test takers' strategy use and test performance on the reading test in CET-4. Phakiti's questionnaire (2008) was employed to measure test takers' strategy use, and the construct of the reading test in CET-4 is lexico-grammatical reading ability (LEX-GR) and text comprehension reading ability (TxtCOM). SEM analysis showed that cognitive strategies had low and non-significant effects on test takers' reading performance. Metacognitive strategies had a significant and direct effect on cognitive strategies, confirming Purpura (1997) and Phakiti's findings (2003). Monitoring strategies were found to significantly influence LEX-GR and evaluating strategies significantly affected TxtCOM. Zhang and Zhang's study (2013) showed a clearer picture of metacognitive strategies on the reading test, confirming Phakiti's finding (2003) that metacognitive strategies had significant effect on reading performance.

The strategy questionnaires in prior studies are primarily based on Purpura's questionnaire embedded in the model of human information processing, which include only cognitive and metacognitive strategies, and might not apply to specific tasks like integrated writing. Hence, Yang and Plakens (2012) developed a strategy use inventory for integrated writing based on theoretical frameworks of integrated writing, empirical

studies on processes of integrated writing and test-taking strategies. The strategies include self-regulatory strategy use, discourse synthesis strategy use, and "test-wiseness" strategy use. They used questionnaire and retrospective interview to study the relationship between strategy use and 161 test takers' performances on integrated writing tasks. The writing ability assessed in the study consists of content, organization, and language use. SEM analysis showed that self-regulatory strategy, as one type of metacognitive strategies, had an executive impact on other strategies, echoing the executive role of metacognitive strategies on cognitive strategies (Purpura, 1997; Phakiti, 2003; Zhang & Zhang, 2013). Discourse synthesis strategy had a direct and positive impact on test performance, and "test-wiseness" strategy had a direct and negative impact on test performance. The finding echoed Song and Cheng's study that only some strategies contributed to the test takers' language performance, and this study confirmed Purpura's finding (1997) that some strategies even had negative impact on test performance. In the previous studies, the generalization of the relationship between strategy use and test performance across samples was not analyzed. To address the niche, Zhang, Goh, and Kunnan (2014) adopted the multi-sample SEM analysis to analyze the relationship between test takers' strategy use and their performance on the reading test in CET-4. The cognitive strategies include initial reading, identifying important information, integrating, inference making, and the metacognitive strategies include planning, evaluation, and monitoring. They found that cognitive and metacognitive strategies functioned in a unitary manner to enhance the Chinese test takers' reading test performance, and this relationship is generalizable across samples. However, they regarded cognitive and metacognitive strategies as a whole, and it remained vague how cognitive and metacognitive strategies as well as their sub-strategies affect test takers' test performance. The strategies in previous studies were measured in terms of frequency count, though most of them found that higher levels of frequency of strategy use did not correspond to better language performance (e.g. Purpura, 1997; Song & Cheng, 2008). What's more, though some researchers in the previous studies speculated that language knowledge might be the main factor to mediating the relationship between strategy use and language performance (e.g., Phakiti, 2003; Purpura, 1997), they did not examine the mechanism of the interaction between language knowledge and strategy use. To address the gap, Cai and Kunnan (2020) developed the Strategy Use Ability Scale (SUAS) that emphasized the efficiency of strategies instead of frequency to elicit test takers' strategy use in accomplishing Nursing English Reading Test (NERT). In the questionnaire, the cognitive strategies are comprehending, memory, and retrieving strategies, while the metacognitive strategies are planning, monitoring, and evaluating strategies. They ran multi-layered moderation analysis to analyze the relationship between 1491 nursing students' strategy use and reading performance, and found that the effect of strategy use ability on nursing English reading performance fluctuated in a down-up-down pattern mediated by students' language knowledge, suggesting that strategy use might not work when the language knowledge is extremely inadequate, while strategies might provide less compensation when language knowledge becomes extremely high. This study adds to the understanding of how language knowledge mediated the effect of strategy use on language performance.

By reviewing the literature, it is found that no agreement has been reached as to the relationship between strategy use and language performance, and some limitations of

the previous studies point to further investigation. Firstly, Purpura's questionnaire based on human information processing dominated the measurement of strategies. Other strategies are excluded like affective strategies and compensation strategies. It remains to be seen whether these strategies will exert impact on test takers' test performance. Secondly, the effect of strategy use on language performance differs across tasks, and only one study touched upon integrated task (Yang & Plakens, 2012). Hence, more integrated tasks can be investigated to inspect if there is any pattern for the mediation of task type on the relationship. Thirdly, prior studies primarily adopted questionnaire and interview methods, few has used think-aloud protocol to elicit the online cognitive processes and strategy use.

Methods

To address the research gaps in the studies of paraphrasing strategies and the relationship between strategy use and language performance, the present study has two purposes. Firstly, the present study seeks to use think-aloud protocols and questionnaire of strategy use inventory to elicit novice ESL writers' online cognitive processes and strategy use of paraphrasing so that it can advance the understanding of the construct of paraphrasing. Secondly, it aims to illuminate the relationship between paraphrasing strategy use and paraphrasing performance so that it can provide insights into this controversial topic in the literature.

Research questions

To this end, two research questions are posed to guide the data collection and analyses:

I. What characterizes test takers' cognitive processes and strategies in responding to the paraphrasing task?

II. What is the relationship between test takers' paraphrasing strategy use and paraphrasing performance?

Cognitive processes and strategies differ in automaticity and purposefulness: the former is more subconscious and habitual, while the latter is more conscious and willful (Chamot, 1987; Cohen, 1998). The two terms are used interchangeably in this paper.

Instruments

Instruments include the paraphrasing task, rating scale for paraphrasing, think-aloud protocol, and paraphrasing strategy use inventory.

Paraphrasing task

Taking a developmental perspective, ESL writers make transition from using paraphrasing as a knowledge telling strategy to using paraphrasing as a knowledge transforming strategy (Currie, 1998). This study focuses on the novice ESL writers, so the task is designed to assess paraphrasing as a knowledge telling strategy, that is to say, to see if they can provide a faithful representation of the original texts. The paraphrasing task provided participants with two reading passages, and participants were asked to paraphrase four and three underlined paraphrasable sentences in passage 1 and passage 2 respectively.

According to Saville-Troike (2008), productive competence follows the receptive competence, which was also echoed by Brown (2002) that production falls behind comprehension, and people understand more language than they can use. A pilot study prior to the present study indicated that as a productive skill, paraphrasing is more demanding than reading, and participants whose English proficiency is at CET-4 level found it demanding to paraphrase sentences in CET-4 reading materials. Therefore, the difficulty level of the reading material was lowered, and the two reading passages in the paraphrasing task were selected from English test papers in College Entrance Examination in China. The criterion of choosing the seven targeted sentences is that they are critical to the comprehension of the whole passages. Thus, there are 7 items in total, and each sentence to be paraphrased constitutes one item. The reading passages were available to participants while they were responding to the paraphrasing task. The difficulty levels of the two passages were examined with Coh-Metrix, and the readability statistics (i.e., narrativity, syntactic simplicity, word concreteness, referential cohesion, deep cohesion, and Flesch Klincaid Grade Level) showed that the difficulty levels of the two passages are comparable (see Figs. 1 and 2).

Rating scale for paraphrasing task

To evaluate participants' paraphrasing performance, an analytic rating scale was developed drawing on Bachman's model of communicative language ability (Bachman, 1990). Organizational and pragmatic competence are involved in paraphrasing task. To be more specific, the subcomponent of organizational competence, that is, grammatical competence, including vocabulary, morphology, and syntax, is reflected in the lexical, morphological, and syntactic transformations of the original texts by substituting words, changing word forms, restructuring sentences, and ensuring grammatical accuracy. The subcomponent of pragmatic competence, that is, sociolinguistic competence is involved in paraphrasing, lying in test takers' sensitivity to register. Since paraphrasing is mainly used in academic setting, language is expected to be formal and explicit.

Aligning the general language proficiency model with paraphrasing, two dimensions are covered in the scale, which are content and language. The former mainly concerns the faithful and full representation of original texts, while the latter deals with the quality of language. Based on the theoretical weighting of the two dimensions, content is assigned 6 points for each of the seven items, thus making the total content score 42,
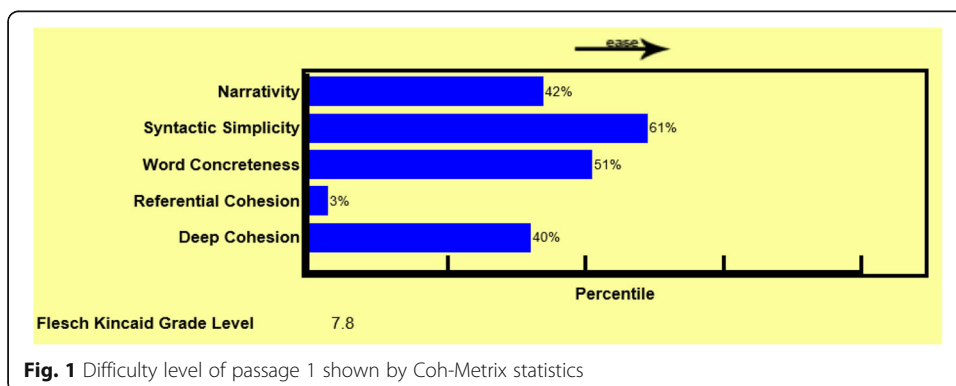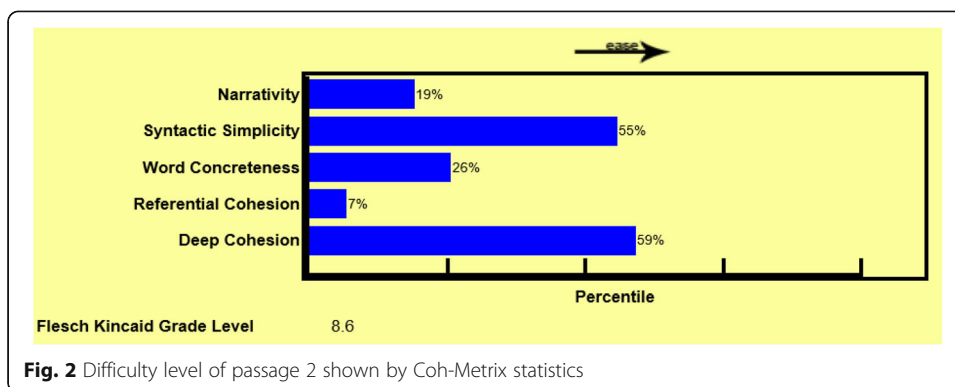

Fig. 1 Difficulty level of passage 1 shown by Coh-Metrix statistics

Fig. 2 Difficulty level of passage 2 shown by Coh-Metrix statistics

and language is given 4 points for each of the seven items, making the total language score 28. To ensure the reliability of the scale, Many-facet Rasch Measurement (MFRM) was performed to inspect the item measurement, scale measurement, and rater severity, and the fit statistics showed that all functioned satisfactorily: the infit values fall within the acceptable range of 0.6 to 1.4 (Linacre, 2005), and ZStd values are acceptable being within − 2 to + 2 (Wright et al., 1994). Due to the space limit, the detailed description of the MFRM results was omitted. Furthermore, expert judgement was employed to check the quality of the scale: three experts in language testing inspected and approved the use of it. Table 1 displays the rating scale in this present study.

Think-aloud protocol

To uncover test takers' online cognitive processes of paraphrasing and make test takers' thinking processes as explicit as possible, the method of think-aloud protocol was

Table 1 Rating scale for paraphrasing

Content

6-A response that fully and faithfully represents the points of the source texts and makes a lot of transformations in vocabulary and syntax.

5-A response that faithfully represents the main points of the source texts and makes most of the transformations in vocabulary and syntax, but some minor points are omitted.

4- A response that covers the main points of the source texts and makes some transformations in vocabulary and syntax, but some points are misrepresented.

3-A response that contains the main points of the source texts, and only replaces words with synonyms, or only changes sentence structures, with one major point omitted.

2- A response that contains only a few relevant points in the source texts and makes few transformations in vocabulary and syntax, and the points are totally misrepresented.

1-A response that fails to present any meaningful or relevant information from the source.

0-A response that is not connected to the topic, written in Chinese, or left blank.

Language

4-A response that changes most expressions in the source texts, is accurately presented, with the participant's own formal and explicit language.

3-A response that changes many expressions in the source texts, with some grammatical mistakes or some complex, implicit and fancy words.

2-A response that only replaces key words with synonyms, and keeps many expressions of the source texts.

1-A response that displays many instances of verbatim strings from the source texts.

0-A response that displays wholesome copying from the source texts.

conducted, which are revealed to have some advantages (Faerch & Kasper, 1987; Green, 1998; Huot, 1993). For example, they are "particularly informative about informants" global approach to a task, the levels of decision-making they operate on, and the considerations that govern their decisions (Faerch & Kasper, 1987, p. 16), and they are immediate, avoiding problems of information retrieval or filtering (Green, 1998). In addition, they are more likely to reflect what participants actually do rather than what they believe they do (Huot, 1993).

As many as 12 participants were recruited to conduct the think-aloud protocol. To explore whether there is some variation in their paraphrasing strategy use across English proficiency levels, they were purposively selected from three levels of English proficiency based on their overall CET-4 scores. Four participants whose CET-4 scores range from 610 to 638 were categorized into the high-proficiency group (M = 623.75, SD = 12.97) and four participants whose CET-4 scores range from 582 to 595 belong to the middle-proficiency group (M = 587.25, SD = 6.40), and the remaining four with CET-4 scores from 547 to 579 belong to the low-proficiency group (M = 568.00, SD = 14.37).

Before conducting the think-aloud protocol, participants were briefly trained. Participants were acquainted with think-aloud guidelines that instruct them to use Chinese or English to articulate their thinking, and not to over explain or analyze (Green, 1998; Perkins, 1981; Xu & Wu, 2012). Their verbalizations were recorded and transcribed in a word-for-word manner for the sake of further coding and analysis. The coding theme of the oral report was developed by borrowing reading strategies (Purpura, 1997), integrated writing strategies (Yang & Plakens, 2012), language learning strategies (O'Malley & Chamot, 1990; Oxford, 1990), and paraphrasing strategies from writing centers of English-medium universities. In total, there are four types of strategies, which are cognitive, metacognitive, compensation, and affective strategies (see Table 2). In the cognitive strategy section, comprehending, repeating, memorizing, retrieving, summarizing, and analyzing are involved. In the metacognitive strategy section, planning, monitoring, and evaluating are included. In the compensation strategy part, guessing and approximating are invoked. Encouraging yourself is regarded as the sole sub-variable of affective strategies. The report of the think-aloud protocol was coded by two coders. They initially coded the dataset independently and then discussed and settled any discrepancy to reach consensus.

Paraphrasing strategy use inventory

Despite the various advantages of think-aloud protocols, they also have several limitations, such as being difficult to administer because test takers may feel uncomfortable in verbalizing their internal thoughts while completing a task (Smagorinsky, 1994). What's more, the process of transcribing, coding, and analyzing data from think-aloud protocols is time-consuming and labor-intensive (Green, 1998; Smagorinsky, 1994). The main criticism of think-aloud protocols is veridicality and reactivity. The former refers to whether participants accurately report their true and complete thinking processes, while the latter is about whether the reporting behavior alters their thinking processes of responding to a task (Ericsson & Simon, 1993; Lumley, 2005; Russo, Johnson, & Stephens, 1989; Stratman & Hamp-Lyons, 1994). Furthermore, the results

Table 2 Coding theme of the think-aloud protocol

Variables
Cognitive strategies
Comprehending
Repeating
Memorizing
Retrieving
Analyzing
Summarizing
Metacognitive strategies
Planning
Monitoring
Evaluating
Compensation strategies
Guessing
Approximating
Affective strategies
Encouraging yourself

generated by think-aloud protocols are individual in nature and may not provide generalized statements (Hwang & Lee, 2017).

Being easy to administer and analyze, eliminating the influence of the verbalizing behavior on the revelation of thinking processes, and gathering data which can generalize to large samples and statistical techniques, a strategy use inventory can serve as a supplementary method to overcome most of disadvantages of think-aloud protocols and help to form a more nuanced picture of test takers' paraphrasing processes. Based on the criteria for developing a valid and reliable inventory (Dornyei, 2003; Gilham, 2000), a 22-item PSUI was developed to capture test takers' mental activities in completing the paraphrasing task, with the composites for the paraphrasing strategy variables being the same as the coding theme of the think-aloud protocol. To establish content validity, one expert and two post-graduate students in applied linguistics were asked to scrutinize the simplicity, clarity, and readability of items. Based on their suggestions, the PSUI were revised. To reduce test takers' cognitive demand while reporting their strategy use, the PSUI written in English was then translated into Chinese by the author.

Like the coding theme of the think-aloud protocol, the PSUI includes cognitive, metacognitive, compensation, and affective strategies as well as sub-strategies (see Table 3). A 6-point Likert scale is used: 0 (never), 1 (very rarely), 2 (rarely), 3 (occasionally), 4 (often), and 5 (very often), which is to instruct participants to indicate their frequency on the use of strategies. The strategy variables have acceptable Cronbach alpha values above 0.500, which are relatively low but according to Hinton et al. (2014), the reliability of 0.500 to 0.75 is "generally accepted to indicate a moderately reliable scale" (p. 363).

Table 3 Composite for paraphrasing strategy use inventory

Variables	No.	Items
Cognitive strategies		
Comprehending	2 items	2, 13
Repeating	2 items	15, 16
Memorizing	1 item	9
Retrieving	2 items	8, 12
Analyzing	2 items	3, 20
Summarizing	2 items	6, 7
Subtotal	11 items	
Metacognitive strategies		
Planning	2 items	4, 19
Monitoring	2 items	5, 14
Evaluating	2 items	21, 22
Subtotal	6 items	
Compensation strategies		
Guessing	2 items	17, 18
Approximating	2 items	10, 11
Subtotal	4 items	
Affective strategies		
Encouraging yourself	1 item	1
Subtotal	1 item	

Procedures

Participants

A total of 212 participants from a key university in the southwest of China were recruited in this study to fill in the consent form, respond to the paraphrasing task, and reveal their cognitive processes and strategy use in paraphrasing. They were first-year non-English majors from college English class, and all of them have learned English as a foreign language for about a decade, and they have passed CET-4, which is the most influential nationwide standardized college English test in China (Jin, 2008), administered by the National College English Testing Committee (Zheng & Cheng, 2008). The mean of their CET-4 overall score is 582.35 (SD = 25.48). As first-year college students, they just started to be engaged with academic writing practices and have not received adequate training in academic writing, so they are defined as novice ESL writers in the present study. Their ages range from 18 to 20 years old (M = 19.23, SD = 0.81), among whom there are 65 females and 147 males. They first read the instructions of paraphrasing and an example of good paraphrasing was presented right after the instructions, and they were then assigned 30 min to complete the task in a pencil-and-paper fashion. Immediately after the paraphrasing test, they were asked to complete the PSUI within 10 min. Among all participants, twelve were purposively chosen to do the think-aloud protocol. Their verbal reports were recorded.

Rating

As for the rating, two experienced raters with master's degrees in applied linguistics were recruited in the scoring of paraphrasing. Prior to rating was a training session

lasting for 40 min, which started by briefly informing raters of the task instructions, rating scale, and example paraphrasing response to familiarize them with the task and rating scale. After that, they practiced rating several paraphrasing scripts followed by a discussion about rating decisions. Once agreement was reached, they started rating operationally. Each sample was assigned scores by raters independently. The two raters' scores were averaged to decide on the final score, and in case of discrepancy of more than three points, a third rater was involved in the rating, whose score was combined with the closer one of the two original scores to determine the final score, while the more discrepant original score was discarded. The inter-rater reliability is acceptable, as is indicated by the Pearson product-moment correlation coefficient (r = 0.891). As many as 85.3% of test takers' total scores had rater discrepancy within three points, so for these scripts, the scores were the average scores of the two raters. The 14.7% problematic scripts were graded by a third rater, whose score and the closer score of the previous two raters were averaged to generate the final scores.

Data collection

Test takers' cognitive processes and strategies were elicited by think-aloud protocols and the PUSI. The oral report of participants was transcribed and then coded by two coders. Test takers' strategy use data was entered into computer and then analyzed to inspect the relationship between test takers' strategy use and their paraphrasing performance.

Data analysis

Descriptive statistics were calculated to examine how test takers performed on the paraphrasing task. The values of the mean, standard deviation, minimum, maximum, skewness, and kurtosis were presented.

In relation to research question I, participants' verbalizations of their thinking processes were recorded, transcribed, and coded. Two coders agreed with each other for 80.6% of their coding results, and settled all discrepancy one by one to reach the consensus. The frequency of strategy use in three proficiency groups of test takers' think-aloud protocol was calculated manually to inspect the deployment of strategies and whether there is any pattern.

In relation to research question II, exploratory factor analysis (EFA) using principal axis extraction was firstly conducted to examine the structure of the hypothesized variables. The Kaiser-Meyer-Olkin value is 0.799, far above the recommended value of .50 (Pett, Lackey, & Sullivan, 2003), indicating that correlations patterns are relatively compact. Bartlett's test of sphericity reached statistical significance (p = .000), suggesting that the correlation matrix is significantly different from an identity matrix (Field, 2013). The two indices showed that factor analysis is appropriate for the data set. The results showed that 12 items were eliminated due to either low loadings or lack of meaningful interpretability.

Preliminary analysis of confirmatory factor analysis (CFA) and SEM were then conducted. As was suggested by Bollen and Long (1993), Kunnan (1998), and Zhu, Raquel, and Aryadoust (2019), there are five stages in SEM analyses. Firstly, the relationships were specified among variables in one measurement model and one structural model based on the theoretical hypothesis and results of EFA. Secondly, model identification

was examined by calculating the difference between the number of known and un-known parameters (degrees of freedom = 32), and the results suggested that the model was over-identified and ideal for SEM analysis. Thirdly, data preparation was conducted by checking the sample size, univariate and multivariate normality, and multicollinear-ity. As for sample size, there are 212 participants in total, and according to Kline (2015), sample size exceeding 200 is regarded as large. Besides, Bentler and Chou (1987) noted that the person-to-parameter ratio should be 5:1, and in this study, there are 24 parameters to be estimated, and the sample size exceeds the minimum require-ment which is 120. Regarding univariate normality, the skewness and kurtosis values fell within – 2 to 2, demonstrating that this assumption was satisfied (Field, 2013). As regards multivariate normality, according to Byrne (2016), the multivariate value repre-sents Mardia's (1970) coefficient of multivariate kurtosis, and the critical ratio of which represents the normalized estimate of multivariate kurtosis. The multivariate kurtosis value is – 4.159, being lower than 3, and its critical ratio is – 1.954, being lower than 5, which all together indicated that the data is multivariate normal. Moreover, multicolli-nearity was checked by VIF value, which was below 5 (ranging from 1.150 to 1.652), implying that it is not a problem in the data (Gujarati & Porter, 2003). Fourthly, max-imum likelihood was used for parameter estimation, since the data is normally distrib-uted. Lastly, multiple fit statistics (i.e., X2/df, TLI, GFI, RMSEA, IFI, PGFI, and PNFI) were used to evaluate the model fit to the data. The results of CFA and SEM from AMOS 23.0 would be reported in the next section.

Results

Descriptive statistics

The result suggested that overall, test takers did not perform quite well (see Table 4), with the average total score being 41.43 out of a maximum of 70 (SD = 10.914), average content score being 24.11 out of a maximum of 42 (SD = 6.791), and average language score being 17.52 out of a maximum of 28 (SD = 4.314). The skewness and kurtosis values are close to 0, thus indicating that the data is normally distributed (Field, 2013).

Think-aloud protocol report

As is shown in Fig. 3, participants indeed employed paraphrasing strategies (N = 216) including cognitive, metacognitive, compensation, and affective strategies. Among the four types of strategies, cognitive strategies were mostly used (N = 122), followed by metacognitive strategies (N = 86), while compensation strategies (N = 3) and affective strategies (N = 5) were less frequently used. Among the three proficiency groups, as their proficiency increases, their use of strategies decreases (N = 85 for low-proficiency

Table 4 Distribution of test takers' total score, content score, and language score

	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Total scores	41.43	10.914	10	67	– 0.193	– 0.137
Content scores	24.11	6.791	5	40	– 0.369	– 0.230
Language scores	17.52	4.314	5	26	– 0.277	– 0.152

The total score of the paraphrasing task is 70; the content score is 42; the language score is 28

Fig. 3 Number of paraphrasing strategies across proficiency groups
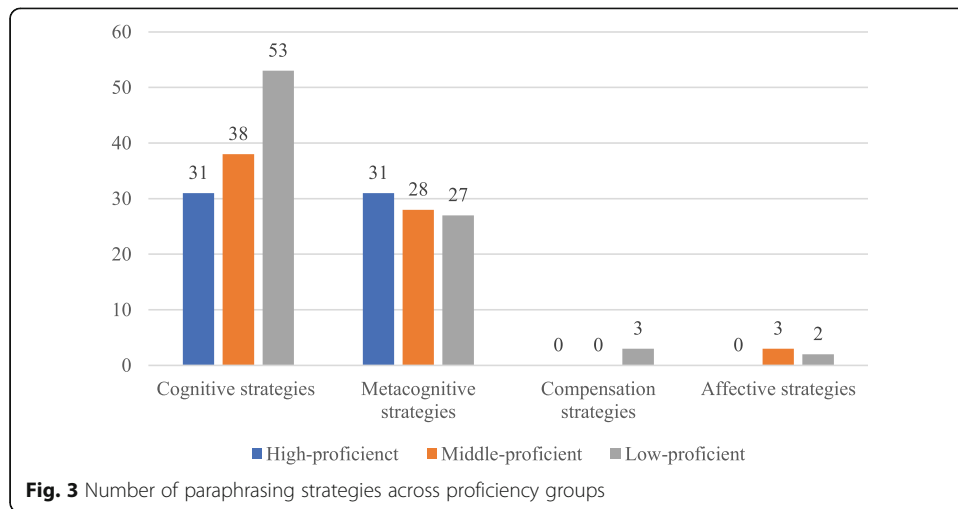
group, $N = 69$ for middle-proficiency group, and $N = 62$ for high-proficiency group). More proficient test takers tend to use more metacognitive strategies.

When it comes to the sub-categories of strategy use (see Fig. 4), as for cognitive strategies, overall, comprehending ($N = 36$), analyzing ($N = 45$), and summarizing strategies ($N = 24$) were employed more than repeating ($N = 10$) and retrieving strategies ($N = 9$) by three proficiency groups. As regards metacognitive strategies, it showed that all the three proficiency groups used monitoring and evaluating strategies a lot ($N = 38$ for monitoring, and $N = 46$ for evaluating). Guessing was used by middle- and low-proficiency participants, while no approximating strategy was used by any three proficiency groups.

EFA results

The EFA analysis showed that items clustered upon three strategy variables: cognitive, metacognitive, and compensation strategies (see Table 5). As for cognitive strategies,
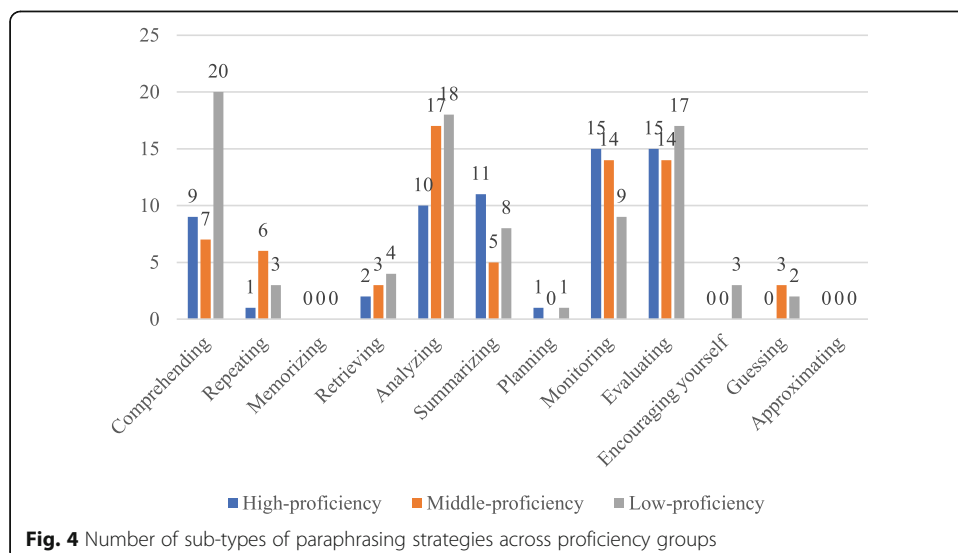

Fig. 4 Number of sub-types of paraphrasing strategies across proficiency groups

Table 5 EFA results of the PSUI: Oblimin Rotation

Item No.	COG	META	COMP
2 (COM)	0.499		
6 (SUM1)	0.627		
7 (SUM2)	0.776		
20 (ANA)	0.804		
5 (MON)		0.584	
21 (EVA1)		0.776	
22 (EVA2)		0.779	
17 (GUE1)			0.845
18 (GUE2)			0.812
11 (APP)			0.435

COM comprehending, *SUM* summarizing, *ANA* analyzing, *REP* repeating, *MON* monitoring, *EVA* evaluating, *GUE* guessing, *APP* approximating, *COG* cognitive strategies, *META* metacognitive strategies, *COMP* compensation strategies

one item of comprehending and two items of summarizing have relatively high loadings. The other item of comprehending was dropped due to low loadings, suggesting that comprehending is not a major strategy in paraphrasing as an integrated task involving both reading and writing. Items tapping into retrieving, memorizing, and repeating were deleted due to low loadings or lack of interpretability, implying that as an integrated task, paraphrasing does not involve as many retrieval, memory, and careful reading as reading. Regarding metacognitive strategies, two items representing planning and one item of monitoring were eliminated due to low loadings or loading on other variables which cannot be meaningfully explained. This indicates that paraphrasing needs little planning like writing, and monitoring is not very much invoked as well. For compensation strategies, one item of approximating is dropped due to low loadings, suggesting that test takers may not regard giving less accurate response as a strategy.

CFA results

The measurement model was used to validate the relationship between latent and observed variables. The measurement model in the present study included cognitive strategies, metacognitive strategies, and compensation strategies (Fig. 5). Affective strategies were deleted after EFA analysis. The hypothesized model showed a misfit for the sample data: the chi-square value being 61.443, the value of degrees of freedom is 32, and p value is 0.00. Based on the modification statistics, a change was made by estimating a covariance parameter between the errors associated with e11 and metacognitive strategies, which can be theoretically explained by the involvement of monitoring and evaluating in using the approximating strategy in that it calls for checking and revising. The modified model was then tested and demonstrated model fit. The chi-square statistic is 44.290 with 31 degrees of freedom (p = .06). Below is the fitting measurement model. Other fit statistics were also displayed below in Table 6, which all provided complimentary evidence that the model is fitting.
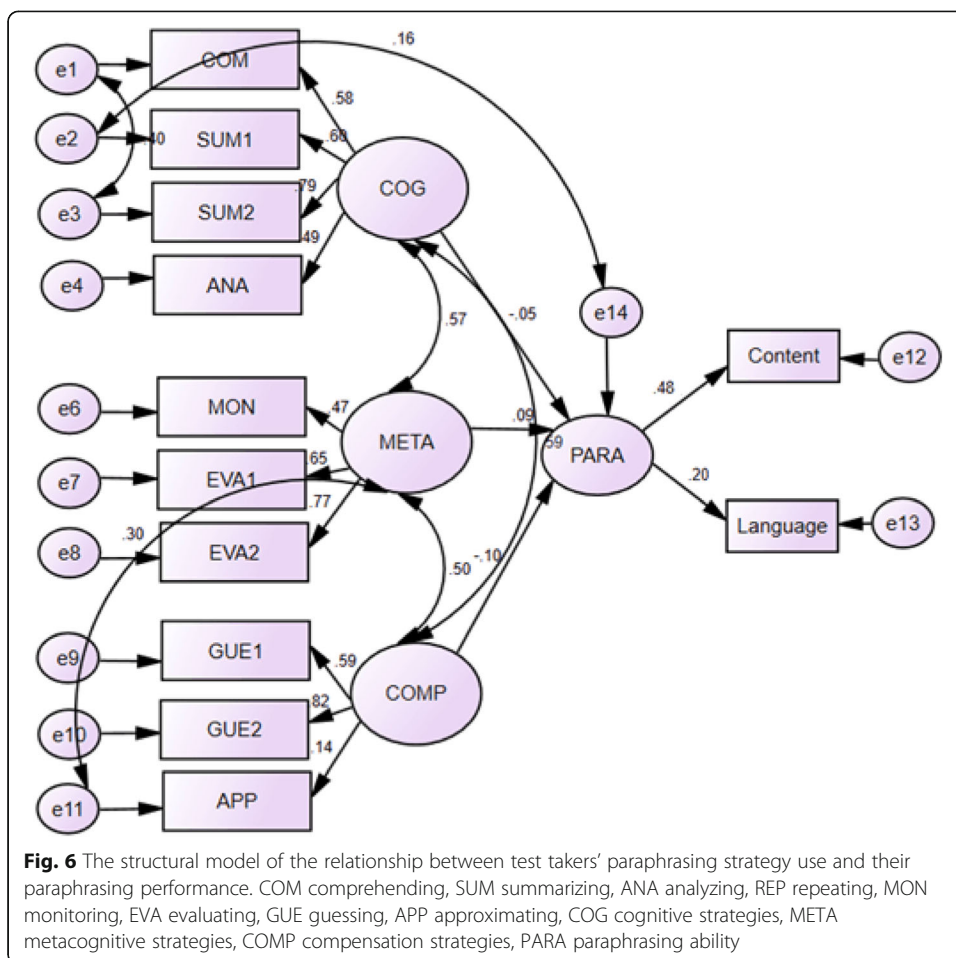
Fig. 5 The measurement model of the relationship between test takers' paraphrasing strategy use and their paraphrasing performance. COM comprehending, SUM summarizing, ANA analyzing, REP repeating, MON monitoring, EVA evaluating, GUE guessing, APP approximating, COG cognitive strategies, META metacognitive strategies, COMP compensation strategies
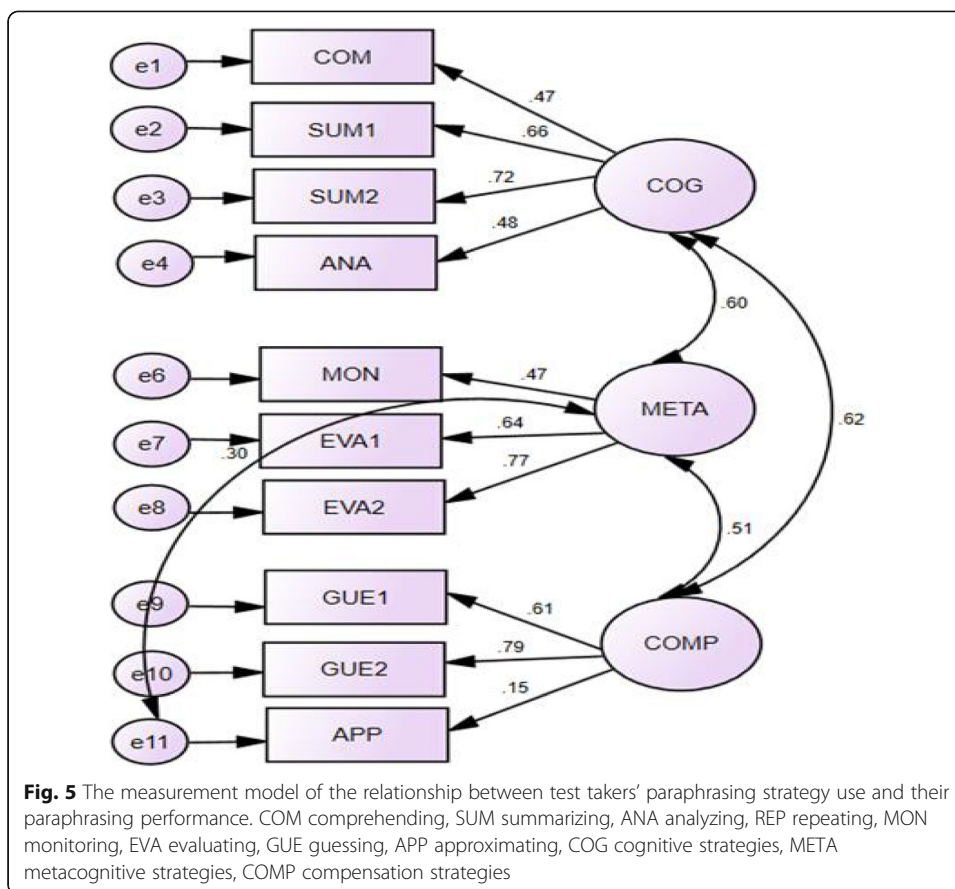
SEM results

SEM was chosen as the primary data analysis tool because research has demonstrated that SEM has numerous advantages over other multivariate procedures. For example, it can correct measurement errors (Jöreskog & Sörbom, 1989; Stevens, 1992; Yang & Plakens, 2012), allow observed and latent variables to be tested simultaneously (Byrne, 2016), and draw a clear map between the latent and observed variables (Zhu et al., 2019). After running CFA, SEM was conducted to model the relationship between test takers' paraphrasing strategy use and their paraphrasing performance. The structural model was found to represent the sample data well (see Fig. 6). The model yielded a chi-square value of 73.578, with 47 degrees of freedom (p = .008), suggesting that the actual model is significantly different from the hypothesized one, though other fit statistics were acceptable. Therefore, modifications were made by freeing the estimate between e2 and e14, e1, and e3, which seems interpretable in that summarizing plays a critical role in paraphrasing, and comprehension of the original texts is helpful for summarizing the main idea of the original texts. The modifications resulted in a fitting

Table 6 Fit statistics of the measurement model

	X²/df	TLI	CFI	RMSEA	IFI	PGFI	PNFI
Expected	≤ 3.0	≥ 0.90	≥ 0.90	≤ 0.10	≥ 0.90	≥ 0.50	≥ 0.50
Observed	1.429	0.952	0.967	0.045	0.968	0.541	0.620

Fig. 6 The structural model of the relationship between test takers' paraphrasing strategy use and their paraphrasing performance. COM comprehending, SUM summarizing, ANA analyzing, REP repeating, MON monitoring, EVA evaluating, GUE guessing, APP approximating, COG cognitive strategies, META metacognitive strategies, COMP compensation strategies, PARA paraphrasing ability

model with a chi-square value of 58.105, with 45 degrees of freedom (p = .091). Other fit indices were shown below (Table 7).

However, the path coefficients in the structural model (see Fig. 6) indicated that the three strategy variables all loaded weakly on the construct of paraphrasing (– 0.05, 0.09, – 0.10 respectively), and the relationship was found to be statistically insignificant in the data shown by the p values in Table 8.

Discussion

Research question I. What characterizes test takers' cognitive processes and strategies in responding to the paraphrasing task?

The data from think-aloud protocols and the PSUI showed that a great many strategies were adopted by participants, including cognitive, metacognitive, compensation and affective strategies with some sub-strategies of cognitive strategies like comprehending, analyzing, and summarizing, and sub-strategies of metacognitive strategies like

Table 7 Fit statistics of the structural model

	X²/df	TLI	CFI	RMSEA	IFI	PGFI	PNFI
Expected	≤ 3.0	≥ 0.90	≥ 0.90	≤ 0.10	≥ 0.90	≥ 0.50	≥ 0.50
Observed	1.291	0.978	0.985	0.037	0.985	0.553	0.639

Table 8 Relationship between test takers' strategy use and paraphrasing performance

	Standardized estimates	S.E.	C.R.	*p*
PARA←COG	− 0.052	1.307	− 0.503	0.615
PARA←META	0.091	1.518	0.921	0.357
PARA←COMP	− 0.097	1.327	− 0.934	0.350

COG cognitive strategies, *META* metacognitive strategies, *COMP* compensation strategies, *PARA* paraphrasing ability

evaluating and monitoring. This confirms that strategy use is an indispensable part in language test performance (Cohen & Upton, 2007; Phakiti, 2003; Plakans, 2008; Purpura, 1997). Meanwhile, it provides insights into refining the model of strategic competence in Bachman (1990) and Bachman and Palmer (1996), who conceptualized strategic competence as assessment, planning, and execution. This understanding is limited to metacognitive components of strategic competence. Previous studies on strategy use and language performance primarily found that cognitive and metacognitive strategies are frequently used by test takers, though some detected the significant effect on test performance (Phakiti, 2003; Song & Cheng, 2008; Yang & Plakens, 2012; Zhang, Goh, & Kunnan , 2014), and others did not (Purpura, 1997; Zhang & Zhang, 2013). The present study adds to the literature by showing that test takers indeed employed a lot of strategies as well as sub-strategies in response to the task, which include not only cognitive and metacognitive strategies, but also compensation and affective strategies transferred from learner strategies in second language acquisition (SLA). It is noteworthy that despite the fact that Phakiti (2003) focused on cognitive and metacognitive strategies, he observed that affective strategies surfaced in the qualitative data, which had the function of easing pressure and anxiety to enhance test performance. He raised the question whether assessing feelings is related to metacognitive strategies, thus being part of strategic competence. The current study offered additional support for his argument. Therefore, the conceptual framework of strategic competence might be informed and refined by both the theories of language learning strategies in SLA and the empirical findings so that it can be more valid. Metacognitive strategies were found to moderately correlate with cognitive strategies, thus providing additional empirical evidence for the executive function of metacognitive strategies over cognitive strategies, which supported the findings in prior studies (Phakiti, 2003; Purpura, 1997; Yang & Plakens, 2012; Zhang & Zhang, 2013).

It is notable that higher-proficiency participants were found to use fewer strategies than lower-proficiency participants, which contradicted the assumption that more proficient learners use strategies more frequently (Green & Oxford, 1995; Jiménez, García, & Pearson, 1996; Oxford & Burry-Stock, 1995). Phakiti's study (2003) found that more successful test takers were reported to use more strategies than less successful ones. One reason for the discrepancy in the present study might be that the sample size for the think-aloud protocols in the present study is much smaller than Phakiti's study (2003), making it hard for the finding in this study to generalize. Another reason is that frequency count might not be a valid instrument to measure strategy use (Oxford, 2017; Cai & Kunnan, 2020). Rather, the efficiency of strategy use (i.e., doing things rightly) is likely to matter more (Grabe & Stoller, 2011; Griffiths & Inceçay, 2016; Oxford, 2017; Cai & Kunnan, 2020). The validity of frequency and efficiency of strategy use warrants further examination. In addition, more proficient test takers were found

to use more metacognitive strategies than less proficient ones, which is consistent with Phakiti's finding (2003). As was revealed by the participants in the interview in Phakiti's study (2003), more proficient test takers were more conscious of how and why they used strategies and which ones worked efficiently for them to deal with the tasks.

Research question II. What is the relationship between test takers' strategy use and their paraphrasing performance?

SEM analysis indicated that cognitive, metacognitive, and compensation strategies produced no significant impact on test takers' paraphrasing performance, though these strategies were indeed employed by participants. The finding corroborated with Purpura' study (1997) and Zhang and Zhang's study (2013) that both cognitive and meta-cognitive strategies had no direct effect on test takers' language performance. However, in Purpura's study, retrieval strategies had a significant and positive effect on grammar ability as one part of the construct tested, and memory strategies had a significant and negative effect on grammar ability. In Zhang and Zhang's study (2013), monitoring strategies had a significant and positive effect on lexico-grammatical reading ability and evaluating strategies had a significant and positive effect on text comprehension reading ability. In this study, no relationship between sub-strategies and language performance was detected. One possible reason is that paraphrasing is an integrated task, including two components of construct, which are content and language. The two components are still integrated and task-specific, making it hard to associate the construct with certain strategies. For example, "content" covers the faithfulness of rephrasing, the types of transformations, and the extent of transformation, while "language" covers the accuracy and appropriacy of the expressions. One type of sub-strategies might not be enough to work for the paraphrasing performance. By contrast, the construct of the test in Purpura's study (1997) includes reading ability and grammar ability, and the construct of the test in Zhang and Zhang's study (2013) is reading ability, which is relatively general and might have a close relation with certain sub-strategies. The finding in the current study contradicted the findings in Phakiti (2003), Song and Cheng (2008), and Zhang, Goh, and Kunnan (2014) that strategy use had a significant effect on test takers' language performance. Though the significant effect was found in these three studies, the effect size was very small. Hence, it is safe to say that strategy use contributes to a small portion of the language performance and the relationship between strategy use and language performance is not that straightforward.

As Bachman (1990) noted, apart from strategies, there might be other factors that account for the test score like language ability, test method effects, personal attributes, and errors of measurement, which might interact with and mediate strategy use as what Cai and Kunnan (2020) found, language knowledge can mediate the effect of strategy use in a down-up-down pattern with the increase of test takers' language knowledge. Test method is also a possible reason for the discrepancy of the findings because research has found that test methods have a sizable influence on test performance (e.g., Bachman & Palmer, 1981, 1982; Shohamy, 1983, 1984). In the studies of the relationship between strategy use and language performance, the language performance was elicited from achievement tests, placement tests, and proficiency tests with diverse test methods, thus making the language performance not comparable. In terms of personal attributes, different test takers are likely to have different proficiency levels, cultural backgrounds, motivation, learning styles, autonomy, aptitude, affect, beliefs, and so

forth. To capture the complexity of the interpersonal variation of strategy use, Oxford (2017) suggested adopting the idiodynamic method (Gregersen, MacIntyre, & Meza, 2014; MacIntyre, 2012), which involves video recording a sample of action from a focal participant and then ask the participant to self-report ratings on some factors of interest and provide explanation for his or her strategy use. As for errors of measurement, the reliability and validity of the measurement of strategy use in the prior studies were not high. Firstly, the definitions and categorization of strategies warrant further examination. Most studies adopted or adapted Purpura's questionnaire (1997), and failed to recognize the importance of contexts in their studies and examine whether the questionnaire is truly applicable to the situations and cultural contexts in their studies. Oxford (2011) argued that cultural adaptations should be made and the reliability and validity of the questionnaire of strategy use should be reassessed in each study and each context. Griffiths (2013) went further to note that researchers need to reject predetermined strategy classification and construct custom-made instruments to align with the characteristics of particular contexts or adopt a data-driven approach to analyze strategy categories. Secondly, the Likertscale-based questionnaire method in eliciting strategies is not without criticisms. Gu, Wen, and Wu (1995) stated that there are ambiguities in the Likert scale survey, and the distinctions between different categories are vague and hard to make, for example "How often is often?" (p. 19). In addition, as Oxford (2017) noted, simple frequency tabulations and strategy categories (e.g., cognitive, metacognitive strategies) cannot reflect the quality of strategy use, learners' contexts, and other background factors, thus missing a lot of important information. Cai and Kunnan (2020) overcame the limitation of frequency counts by designing a strategy use ability questionnaire which highlights the efficiency of strategy use. To present a more meaningful picture, data of other background factors should be gathered through other methods like interview, thus pointing to the importance of adopting mixed-method approach in studying strategy use and the effect on language performance. Furthermore, Griffiths and Inceçay (2016) mentioned that Likert scales are by nature ordinal and do not generate numerical data. However, most studies of strategy use performed parametric tests such as Pearson product-moment test of correlation, t tests, ANOVAs, and SEM to analyze means, which are not the correct analysis procedures despite the fact that results produced by non-parametric tests do not differ widely from those by parametric test. Strictly speaking, the data produced by Likert scales should be analyzed by non-parametric tests like Spearman's rho, Man-Whitney U, or Kruskall-Wallis (Jamieson, 2004).

Conclusion

The present study presents a complete picture of test takers' processes and strategies of paraphrasing, and it is found that test takers indeed used many paraphrasing strategies including cognitive, metacognitive, compensation, and affective strategies. Cognitive strategies including comprehending, analyzing, and summarizing strategies were most frequently used, followed by metacognitive strategies including evaluating and monitoring strategies. The positive and significant correlation between cognitive and metacognitive strategies lent support to the notion that metacognitive strategies exert an executive effect on cognitive strategies. More proficient test takers tend to use a smaller number of strategies in that language knowledge might be adequate in dealing with the

task and make it unnecessary to use strategies. More proficient test takers used more metacognitive strategies as they are more autonomous learners and might be more aware of how and why they used strategies. However, SEM analysis reported that all those strategies exert no significant effect on paraphrasing performance, which contradicts many previous findings. The discrepancy may be attributed to the small effect of strategies on language performance. Other possible factors contributing to the results were examined including language knowledge, test methods, personal attributes, and errors of measurement.

However, admittedly, this study has some limitations which warrant further study. The first limitation concerns the small sample size, and because of it, the multigroup invariance of the relationship between strategy use and language performance cannot be achieved, and future study can be conducted to examine the difference of the structural model between higher- and lower-proficiency test takers. Secondly, the paraphrasing strategy use inventory suffers from some limitations. Apart from the drawbacks of Likert scale-based questionnaire, for one thing, the questionnaire in the present is not well piloted, leading to the elimination of about half items, so more attention should be paid to the piloting of the instrument in the future study. For another, back translation was not conducted so that the equivalency of the two versions of the PSUI is not ensured. In the future study, more qualitative methods are recommended to be adopted to complement the data analysis, such as case study and narrative study. Thirdly, this test was conducted in a low-stakes assessment context, and the performance outcomes of the test had little relevance to test takers' school achievement; thus, the task motivation was not guaranteed, which proved to significantly influence the task engagement and subsequent achievement (Harackiewicz et al., 2008). The lack of task motivation may threaten the validity of the assessment. Last but not the least, factors affecting the relationship between strategy use and language performance need to be probed into more deeply, thus calling for further studies focusing on the mechanism of how strategy use influences language performance. Longitudinal studies are recommended to track the effect of strategy use on language performance at different developmental levels.

Despite the aforementioned limitations, the present study has important implications. Firstly, a clear picture of paraphrasing strategies has been provided, which advances the understanding of paraphrasing construct and paraphrasing development. This helps to guide the instruction and assessment of paraphrasing so that strategy instruction can be adjusted based on ESL writers' developmental stages. Secondly, the current study found that test takers employed not only cognitive and metacognitive strategies but also compensation and affective strategies, which helps to refine the metacognition-oriented model of strategic competence in Bachman (1990) and Bachman and Palmer (1996). Thirdly, the insignificant relationship between paraphrasing strategy use and performance revealed that the overwhelming research method, that is, Likert scale-based questionnaire might have some problems. It might contribute to the mixed results in the literature. The use of think-aloud protocols in the present study showed that qualitative research methods have a complementary role to play in the study of the relationship between strategy use and language performance, and in future studies, more qualitative methods can be conducted such as think-aloud protocols, interview, and narrative study.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s40468-020-00104-3.

Additional file 1.

References

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing, 19*, 453–476.

Bachman, L. F., & Palmer, A. S. (1981). The construct validation of the FSI oral interview. *Language Learning, 31*, 67–86.

Bachman, L. F., & Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly, 16*, 449–465.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Bentler, P.M., & Chou, C. (1987). Practical issues in structural modeling. *Sociological Methods and Research, 16*, 78–117.

Bollen, K.A., & Long, J.S. (1993). Testing structural equation models. Newbury Park, CA: Sage.

Bridgeman, B., & Carlson, S. (1983). Survey of academic writing tasks required of graduate and undergraduate foreign students (Research Report No. RR–83–18). Retrieved from Wiley Online Library website: https://onlinelibrary.wiley.com/doi/abs/10.1002/j.2330-8516.1983.tb00018.x

Brown, H. D. (2002). *Principles of language learning and teaching*. Beijing: Foreign Language Teaching and Research Press.

Burstein, J., Flor, M., Tetreault, J., Madnani, N., & Holtzman, S. (2012). Examining linguistic characteristics of paraphrase in test-taker summaries (Research Report No. RR–12–18). Retrieved from ETS Research website: https://www.ets.org/Media/Research/pdf/RR-12-18.pdf

Byrne, B. M. (2016). *Structural equation modeling with EQS: Basis concepts, applications, and programming*. Mahwah, NJ: Psychology Press.

Cai, Y., & Kunnan, A.J. (2020). Mapping the fluctuating effect of strategy use ability on English reading performance for nursing students: A multi-layered moderation analysis approach. *Language Testing, 37*, 280–304.

Campbell, C. (1990). Writing with others' words: Using background reading text in academic compositions. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 211–230). Cambridge: Cambridge University Press.

Chamot, A. U. (1987). The learning strategies of ESL students. In A. L. Wenden & J. Rubin (Eds.), *Learner Strategies in Language Learning* (pp. 71–84). Englewood Cliffs, NJ: Prentice Hall.

Cohen, A. D. (1998). Strategies and processes in test taking and SLA. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research*. Cambridge: Cambridge University Press.

Cohen, A. D., & Upton, T. A. (2007). "I want to go back to the text": Response strategies on the reading subtest of the new TOEFL. *Language Testing, 24*, 209–250.

Currie, P. (1998). Staying out of trouble: Apparent plagiarism and academic survival. *Journal of Second Language Writing, 7*, 1–18.

Dornyei, Z. (2003). *Questionnaires in second language research: Construction, administration, and processing*. New Jersey: Lawrence Erlbaum Associates.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. London: The MIT Press.

Faerch, C., & Kasper, G. (1987). *Strategies in interlanguage communication*. London: Longman.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage Publications.

Gebril, A. (2010). Bringing reading-to-write and writing-only assessment tasks together: A generalizability analysis. *Assessing Writing, 15*, 100–117.

Gilham, B. (2000). *Developing a questionnaire*. London: Continuum.

Grabe, W., & Stoller, F. L. (2011). *Teaching and researching reading*. London: Pearson.

Green, A. (1998). *Verbal protocol analysis in language testing research: A handbook*. Cambridge: Cambridge University Press.

Green, N. M., & Oxford, R. (1995). A closer look at learning strategies, L2 proficiency, and gender. *TESOL Quarterly, 29*, 261–297.

Gregersen, T., MacIntyre, P. D., & Meza, M. D. (2014). The motion of emotion: Idiodynamic case studies of learners' foreign language anxiety. *Modern Language Journal, 98*, 574–588.

Griffiths, C. (2013). *The strategy factor in successful language learning*. Bristol, UK: Multilingual Matters.

Griffiths, C., & Inceçay, G. (2016). New directions in language learning strategy research: Engaging with the complexity of strategy use. In C. Gkonou, D. Tatzl, & S. Mercer (Eds.), *New directions in language learning psychology* (pp. 25–38). New York: Springer.

Gu, Y., Wen, Q., & Wu, D. (1995). How often is Often? Reference ambiguities of the Likert-scale in language learning strategy research. Occasional Papers in English Language Teaching, 5, 19–35.

Gujarati, D., & Porter, D. (2003). Multicollinearity: What happens if the regressors are correlated. *Basic Econometrics*, 363–364.

Hale, G., Taylor, C., Bridgeman, B., Carson, J., Kroll, B., & Kantor, R. (1996). *A study of writing tasks assigned in academic degree programs*. Princeton: Educational Testing Service.

Harackiewicz, J. M., Durik, A. M., Barron, K. E., Linnenbrink-Garcia, L., & Tauer, J. M. (2008). The role of achievement goals in the development of interest: Reciprocal relations between achievement goals, interest, and performance. *Journal of Educational Psychology, 100*, 105–122.

Hinton, P. R., Brownlow, C., McMurray, R., & Cozens, B. (2014). *SPSS explained*. New York, NY: Routledge.

Hirvela, A., & Du, Q. (2013). "Why am I paraphrasing?": Undergraduate ESL writers' engagement with source-based academic writing and reading. *Journal of English for Academic Purposes, 12*, 87–98.

Howard, R. (1996). Plagiarisms, authorships, and the academic death penalty. *College English, 57*, 788–806.

Huot, B. A. (1993). The influence of holistic scoring procedures on reading and rating student essays. In M. M. Williamson & B. A. Huot (Eds.), *Validating holistic scoring for writing assessment: Theoretical and empirical foundations* (pp. 206–236). Creskill, NJ: Hampton Press.

Hwang, M., & Lee, H. K. (2017). Development and validation of the English writing strategy inventory. *System, 68*, 60–71.

Hyland, F. (2001). Dealing with plagiarism when giving feedback. *ELT Journal, 35*, 375–381.

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education, 38*, 1212–1218.

Jiménez, R. T., García, G. E., & Pearson, P. D. (1996). The reading strategies of bilingual Latina/o students who are successful English readers: Opportunities and obstacles. *Reading Research Quarterly, 31*, 90–112.

Jin, Y. (2008). Powerful tests, powerfulness test designers? Changes facing the College English Test. *CELEA Journal, 31*, 3–11.

Johns, A. M., & Mayes, P. (1990). An analysis of summary protocols of university ESL students. *Applied Linguistics, 11*, 253–271.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications*. Chicago, IL: SPSS.

Keck, C. (2006). The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *Journal of Second Language Writing, 15*, 261–278.

Kucer, S. L. (1985). The making of meaning: Reading and writing as parallel processes. *Written Communication, 2*, 317–336.

Kunnan, A. J. (1998). An introduction to structural equation modelling for language assessment research. *Language Testing, 15*, 295–332.

Linacre, J. M. (2005). *A user's guide to FACETS: Rasch-Model Computer Program*. Chicago: MESA Press.

Lumley, T. (2005). *Assessing second language writing: The raters' perspective (Unpublished doctoral dissertation)*. Melbourne: The University of Melbourne.

MacIntyre, P. D. (2012). The idiodynamic method: A closer look at the dynamics of communication traits. *Communication Research Reports, 29*, 361–367.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*, 519–530.

McGinley, W. (1992). The role of reading and writing while composing from sources. *Reading Research Quarterly, 27*, 227–248.

Mori, M. (2019). Pluralizing the paraphrase: Reconsidering meaning via a typology of linguistic changes. *TESOL Quarterly, 53*, 885–895.

O'Malley, M. J., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.

Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Boston: Heinle & Heinle.

Oxford, R. L. (2011). *Teaching and researching language learning strategies*. Harlow, UK: Pearson Education.

Oxford, R. (2017). *Teaching and researching language learning strategies: Self-regulation in context*. New York, NY: Routledge.

Oxford, R. L., & Burry-Stock, J. (1995). Assessing the use of language strategies worldwide with the ESL/EFL version of the Strategy Inventory for Language Learning (SILL). *System, 23*, 1–23.

Pennycook, A. (1996). Borrowing other's words: Texts, ownership, memory, and plagiarism. *TESOL Quarterly, 30*, 210–230.

Perkins, D. N. (1981). *The mind's best work*. Cambridge: Harvard University Press.

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Thousand Oaks, CA: Sage.

Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing, 20*, 26–56.

Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing, 13*, 111–129.

Purpura, J. E. (1997). An analysis of the relationships between test takers' cognitive and metacognitive strategy use and second language test performance. *Language Learning, 47*, 289–325.

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition, 17*, 759–769.

Saville-Troike, M. (2008). *Introducing second language acquisition*. Cambridge: Cambridge University Press.

Shi, L. (2010). Textual appropriation and citing behaviors of university undergraduates. *Applied Linguistics, 31*, 1–24.

Shi, L. (2016). Textual borrowing in second-language writing. *Written Communication, 21*, 171–200.

Shi, L., & Dong, Y. N. (2018). Chinese graduate students paraphrasing in English and Chinese contexts. *Journal of English for Academic Purposes, 34*, 46–56.

Shohamy, E. (1983). The stability of oral proficiency assessment on the oral interview testing procedure. *Language Learning, 33*, 527–540.

Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing, 1*, 147–170.

Smagorinsky, P. (1994). *Speaking about writing*. London: Sage Publications.

Song, X., & Cheng, L. (2008). Language learner strategy use and test performance of Chinese learners of English. *Language Assessment Quarterly, 3*, 243–266.

Spivey, N. (1990). Transforming texts: Constructive processes in reading and writing. *Written Communication, 7*, 256–287.

Stevens, J. (1992). *Applied multivariate statistics for the social science*. Hillsdale, NT: Lawrence Erlbaum.

Stratman, J. F., & Hamp-Lyons, L. (1994). Reactivity in concurrent think-aloud protocols. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 89–112). London: Sage Publications.

Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing, 9*, 27–55.

Wright, B. D., Linacre, J. M., Gustafson, J. E., & Martin-Lof, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*, 370.

Xu, Y., & Wu, Z. M. (2012). Test-taking strategies for a high-stakes writing test: An exploratory study of 12 Chinese EFL learners. *Assessing Writing, 17*, 174–190.

Yang, H. C., & Plakens, L. (2012). Second language writers' strategy use and performance on an integrated reading-listening-writing task. *TESOL Quarterly, 46*, 80–103.

Zhang, L., & Zhang, L. G. (2013). Relationships between Chinese college test takers' strategy use and EFL reading test performance: A structural equation modeling approach. *RELC Journal, 44*, 35–57.

Zhang, L., Goh, C. C. M., & Kunnan, A. J. (2014). Analysis of test takers' metacognitive and cognitive strategy use and EFL reading test performance: A multi-sample SEM approach. *Language Assessment Quarterly, 11*, 76–102.

Zheng, Y., & Cheng, L. (2008). Test review: College English Test (CET) in China. *Language Testing, 25*, 408–417.

Zhu, X., Raquel, M., & Aryadoust, V. (2019). Structural equation modeling to predict performance in English proficiency tests. In V. Aryadoust & M. Raquel (Eds.), *Quantitative data analysis for language assessment volume II: Advanced methods* (pp. 101–126). London: Routledge.

Publisher's Note