

Profiling L2 learners' paraphrasing competence in English academic writing

Emily Di Zhang¹, Nan Zhou²; Shulin Yu³

School of Foreign Languages, Shanghai Jiao Tong University¹; Faculty of Education, University of Macau^{2,3}

YB97712@um.edu.mo; nanzhoumidfrost@um.edu.mo;

shulinyu@um.edu.mo

Abstract

While the importance of paraphrasing in English academic writing is reiterated, empirical research on L2 learners' paraphrasing competence remains scarce. This study utilized a self-developed and validated L2 learner paraphrasing competence scale to explore L2 learners' paraphrasing levels and profiles. A total of 297 university students from China and Japan undertook the questionnaire. Latent profile analysis was conducted on the 297 students' self-reported responses and three distinct profiles of L2 learner paraphrasing competence were identified: struggling paraphrasers (SP), average paraphrasers with active cognitive processes (APACP), and capable paraphrasers with strong lexical paraphrasing competence (CPSLP). The Bolck-Croon-Hagenaars (BCH) analyses indicated that participants' country, grade, and L2 proficiency were significantly associated with profile membership, while participants' major showed no significant association. The findings shed new light on the complexity and multidimensionality nature of L2 learners' paraphrasing in the East Asian context.

Keywords: paraphrasing, L2 learners, East Asian context, latent profile analysis, higher education.

Resumen

Perfil de la competencia de parafraseo de estudiantes de L2 en la escritura académica en inglés

Aunque la importancia del parafraseo en la escritura académica en inglés ha sido reiteradamente destacada, la investigación empírica sobre la competencia de parafraseo de los estudiantes de L2 sigue siendo limitada. Este estudio utilizó una escala de competencia de parafraseo para estudiantes de L2, desarrollada y

validada por los autores, con el fin de explorar los niveles y perfiles de competencia de los estudiantes en esta habilidad. Un total de 297 estudiantes universitarios de China y Japón respondieron al cuestionario. Se llevó a cabo un análisis de perfiles latentes a partir de las respuestas autoinformadas de los 297 participantes y se identificaron tres perfiles diferenciados de competencia de parafraseo en estudiantes de L2: parafraseadores con dificultades (SP), parafraseadores de nivel medio con procesos cognitivos activos (APACP) y parafraseadores competentes con una sólida competencia de parafraseo léxico (CPSLP). Los análisis de Bolck-Croon-Hagenaars (BCH) indicaron que el país de procedencia, el curso académico y la competencia en L2 de los participantes estaban significativamente asociados con la pertenencia a los distintos perfiles, mientras que la especialidad académica no mostró una asociación significativa. Los resultados aportan nuevas perspectivas sobre la complejidad y la naturaleza multidimensional del parafraseo en estudiantes de L2 en el contexto de Asia Oriental.

Palabras clave: parafraseo, estudiantes de L2, contexto de Asia Oriental, análisis de perfiles latentes, educación superior.

1. Introduction

As English academic writing has drawn increasing attention from researchers and teaching practitioners, there is growing evidence that paraphrasing closely relates to students' academic writing performance in higher education (Du & Liu, 2024; Xu & Zheng, 2025). As a source use strategy, paraphrasing refers to the faithful reformulation of the source texts with acknowledgement to the authors (Mori, 2019). Compared with other source use strategies, such as quotation and summarization, paraphrasing is more complex and demanding because it involves comprehension, interpretation, and evaluation of the source texts (Keck, 2010; Wette, 2017; Zhang & Yu, 2023).

Although recent evidence points to the importance of L2 paraphrasing, prior research on paraphrasing in academic writing has predominantly examined the revision extent of source texts (Keck, 2006; Shi, 2004), linguistic changes (Burstein et al., 2012; Mori, 2019), purposes (Du & Liu, 2024; Hirvela & Du, 2012) and strategies of paraphrasing (Shi & Dong, 2018). Little information is available about the current level and state of L2 learners' paraphrasing competence in different instructional contexts. Thus, more knowledge is needed regarding L2 students' paraphrasing competence

in L2 academic writing for teaching and learning, and assessment purposes. Furthermore, many existing studies assess paraphrase quality by measuring how much text an L2 learner alters from the original source. This 'revision rate' is then compared to the standard set by native academic writers. This method is problematic because L1 and L2 writers are heterogeneous groups with differing developmental trajectories in paraphrasing (Du & Liu, 2024; Keck, 2014). A paradigm shift is called for from the comparison of L1 versus L2 paraphrasing writers to the delineation of the complexity of L2 writers' paraphrasing to better understand L2 writers' level of paraphrasing competence and needs in paraphrasing development. Earlier studies have primarily adopted a variable-centered method to examine the relationships between paraphrasing competence level and relevant factors at the entire sample level, such as country, grade, major, and L2 proficiency (e.g., Hiverla & Du, 2013; Keck, 2006; Shi, 2004; Shi & Dong, 2018; Zhang & Yu, 2023). They failed to focus on the descriptions of the heterogeneity of L2 learner paraphrasing competence and, specifically, the potential patterns at different levels of L2 learner paraphrasing competence.

To address these research gaps, the present study aims to adopt a person-centered approach to counteract the limitations of the variable-based method approach for the study of L2 learners' paraphrasing competence level and profiles. More specifically, it uses latent profile analysis (LPA) in order to examine the potential heterogeneity among East Asian (Chinese and Japanese) L2 university students' paraphrasing competence patterns. It aimed to explore the potential profiles of L2 university students' paraphrasing competence based on (a) the dimensions of a self-developed and validated paraphrasing competence scale and (b) the associations between paraphrasing competence profiles and country, grade, major, and L2 proficiency.

This study moves beyond deficit-based, native-speaker comparisons to profile L2 paraphrasing competence from a person-centered perspective. Using LPA on survey data from 297 Chinese and Japanese university students, it identifies three distinct learner profiles based on a validated scale. These profiles are significantly associated with country, grade, and L2 proficiency, but not with major, highlighting the complex, multidimensional nature of paraphrasing development in East Asian contexts and offering a refined framework for targeted pedagogy.

2. Literature review

Research on L2 paraphrasing has evolved along two primary lines: defining its multifaceted nature and identifying factors that influence it. First, it establishes paraphrasing as a complex construct encompassing Lexical Paraphrasing, Syntactic Paraphrasing, Plagiarism Avoidance, Source Use, and Cognitive Processes. Second, it reviews the variable-centered research that has identified key factors, such as country, grade, major, and L2 proficiency, influencing this competence.

2.1. Construct of paraphrasing

Paraphrasing is a multidimensional construct, with research delineating its key dimensions to better understand its nature and assessment. This section outlines the five primary dimensions, including Lexical Paraphrasing, Syntactic Paraphrasing, Plagiarism Avoidance, Source Use, and Cognitive Processes.

Lexical Paraphrasing involves changes at the word level, including strategies such as synonym substitution, antonym substitution, morphological changes, and the use of multiple word units (Burstein et al., 2012; Culicover, 1968; Martin, 1976; Mel'nik, 1992; Mori, 2019). Similarly, Syntactic Paraphrasing pertains to transformations at the syntax level, such as shifts in active/passive voice shift, coordination (e.g., turning complex sentences into compound sentences), subordination (e.g., turning main clauses into dependent clauses), relative clause/phrase shift, and noun phrase/pronoun shift (Burstein et al., 2012; Martin, 1976; Mel'nik, 1992). Plagiarism Avoidance constitutes a third key dimension, focusing on the degree of substantial textual transformation. Since a primary function of paraphrasing is to avoid plagiarism (Campbell, 1990), the quality of linguistic transformation is critical. Insufficient or superficial changes to source texts—often termed “patchwriting”—carry a high risk of being flagged as plagiarism (Keck, 2006; Shi, 2004). Effective paraphrasing, therefore, requires substantial reworking of both vocabulary and syntax. While Lexical Paraphrasing, Syntactic Paraphrasing, and Plagiarism Avoidance are key dimensions of knowledge-telling, i.e., the faithful restatement of source material, legitimate paraphrasing competence also encompasses knowledge-transforming, which involves integrating, interpreting, and recontextualizing the source content to advance the writer's own argument (Hirvela & Du, 2013).

Source Use and Cognitive Processes embody knowledge-transforming paraphrasing. In academic writing, paraphrasing is not merely reproduction but a strategic act of source use, i.e., selecting and interpreting relevant source information, which is then deliberately woven into the writer's argumentation to actively construct new knowledge (Shi & Dong, 2018; Sun & Yang, 2015; Wette, 2017; Yamada, 2003). As a high-order skill, paraphrasing involves a recursive series of cognitive processes: comprehending the meaning of the original texts, analyzing the importance of ideas and logic of the original texts, synthesizing the relevant and useful ideas from the source texts, applying linguistic knowledge to write paraphrases, monitoring the paraphrasing processes, and evaluating the quality of the final paraphrases (Shi & Dong, 2018; Shi et al., 2018; Wette, 2017; Zhang, 2020).

2.2. Variables affecting L2 students' paraphrasing competence

The bulk of prior research has adopted a variable-centered approach, examining how specific factors affect paraphrasing competence. In terms of the variable of linguistic and cultural background, typically measured by country, much of the existing research, adopting a contrastive approach, has focused on directly comparing the paraphrasing performance of L1 and L2 writers. This line of inquiry has consistently found notable differences rooted in the participants' country. A central finding is that L2 writers tend to engage in more verbatim source use and make slighter textual revisions than their L1 counterparts (Burstein et al., 2012; Keck, 2006, 2014; Shi, 2004). Furthermore, studies indicate that L2 students are more likely to incorporate source texts without references or to use implicit attributions compared to L1 writers (Moore, 1997; Shi, 2004). This contrastive approach has therefore established a clear paraphrasing competence gap, often attributing the challenges faced by L2 writers to their different linguistic and cultural starting points.

Researchers have also probed into the effect of explicit teaching, typically associated with grade levels, on L2 learner paraphrasing competence. Shi (2004) and Keck (2006, 2014) showed that L2 students relied on much copying or minimal text revision, which can be attributed to the lack of explicit teaching in source use: Chinese L2 teachers emphasized the knowledge-telling dimension of paraphrasing, and thereby Chinese L2 undergraduate students viewed paraphrasing as a decontextualized and linguistic-oriented exercise and struggled with using paraphrasing as a source

use device (Hirvela & Du, 2013). Chinese L2 undergraduate students modeled the source use practice in L1 writing to make sense of source use expectations in L2 academic writing, that is, elevating writing style instead of argumentation (Du & Liu, 2024). In a similar vein, Rinnert and Kobayashi (2005) found that the explicit teaching of academic writing in English in Japanese universities is scant, and Yoshimura (2015) reported that very few Japanese university undergraduate students learned about paraphrasing or received any instruction on how to paraphrase, be it knowledge-telling or knowledge-transforming paraphrasing. Research indicates that Chinese L2 graduate students, who typically received more explicit teaching of academic writing than undergraduates, developed more advanced paraphrasing competence. They demonstrated an ability to select and interpret source information, and to add or extend ideas (Shi & Dong, 2018). Furthermore, they could smoothly integrate sources into their own writing through syntactic reformation and relevant recounting (Shi et al., 2018). Wette (2010) looked into the effect of instructional interventions on L2 undergraduate students' source use competence in the context of New Zealand and revealed that students with explicit teaching of source use displayed an increase in accurate paraphrases and summaries, along with a reduction of direct copying and near copying.

Apart from explicit teaching, Shi et al. (2018) have explored the effect of the variable of major on Chinese L2 learners' paraphrasing competence, and they found that there were no significant differences between the hard and soft disciplines in linguistic adaptations, i.e., knowledge-telling paraphrases, and there was a significant difference in content recontextualization, i.e., knowledge-transforming paraphrases, such that compared with students in hard disciplines, students in soft disciplines focused less on adding their own interpretations. Shi et al. (2018) attributed the difference to the different focuses of hard and soft disciplines: the hard disciplines foreground the research results and findings, and the soft ones highlight the cited researchers and their opinions.

Researchers also took the variable-centered approach to examine the relationship between L2 proficiency and paraphrasing competence. Basham and Rounds (1984) observed that the lack of L2 proficiency in deleting and adding modals, changing verb tense and using adverbs caused students to alter the tone and intent of original texts. Likewise, Currie (1998) reported that inadequate L2 proficiency made students rely on copying as a composing strategy, which was corroborated by Shi (2004), Keck (2006) and

Keck (2014). Zhang and Yu (2023) compared the frequency of linguistic changes in paraphrasing across three English L2 proficiency levels (low, intermediate, and high), and revealed that the use of synonym substitution, morphology, multiple word units, phrase/clause shift, active/passive shift, and conceptual transformation increased in frequency as L2 students' proficiency levels advanced.

While prior research has predominantly employed a variable-centered approach to study L2 learner paraphrasing competence, there is a paucity of studies taking a person-centered approach to classify the entire sample into subgroups and examine the potential patterns configured within the L2 learners' paraphrasing competence. Additionally, while most prior studies have focused on learners in the Chinese tertiary context, they have largely failed to include other L2 contexts and to examine potential differences in paraphrasing competence across these settings. To fill the gap, the present study took a person-centered approach to profile Chinese and Japanese university students' paraphrasing competence and examined the associations between L2 learner paraphrasing competence profiles and the variables of country, grade, major, and L2 proficiency.

Therefore, the three research questions to be addressed are:

1. What are the profiles of Chinese and Japanese L2 learner paraphrasing competence?
2. How do Chinese and Japanese L2 learner paraphrasing competence profiles differ across multiple demographic variables (country, grade, and major)?
3. What are the associations between Chinese and Japanese learners' L2 proficiency and paraphrasing competence profiles?

3. Methods

3.1. Participants and procedures

Participants were recruited from Japan and China. Invitation letters for the current study were sent to the ten college English instructors, and they then shared the invitation letters with their students. As many as 297 participants signed the consent form and voluntarily agreed to participate in the present study, including 232 Chinese university student participants and 65 Japanese

university student participants. Participants' majors covered both hard disciplines ($N = 214$), like computer science, civil engineering and automation, and soft disciplines ($N = 83$) like linguistics, literature and education. There were 100 first-year undergraduate students, 116 second-year undergraduate students, 16 third-year undergraduate students, 31 fourth-year undergraduate students and 34 graduate students.

The participants then completed the survey through the *Wenjuanxing* Platform during the class breaks under the supervision of the teacher. The survey included the participants' country, grade, major, an L2 learner paraphrasing competence scale, and an L2 proficiency self-assessment scale. Participants were informed that they could withdraw from the survey at any time.

3.2. Measures

The study utilized three primary categories of measures to assess the variables of interest: a scale for paraphrasing competence, a set of demographic variables (country, grade, and major), and a self-report measure for L2 proficiency.

3.2.1. L2 learner paraphrasing competence

The L2 learner paraphrasing competence scale was used to assess the participants' current levels of paraphrasing. The scale has a total of 22 items scored on a 5-point scale and is divided into 5 subscales, which are:

- (a) Lexical Paraphrasing (e.g., synonym substitution, multiple word units, and morphology),
- (b) Syntactic Paraphrasing (e.g., noun phrase/pronoun shift, declarative/question shift, and relative clause-noun phrase),
- (c) Plagiarism Avoidance (e.g., reformulation and avoidance of copying),
- (d) Source Use (e.g., adding new ideas, critical evaluation, and giving citations) and
- (e) Cognitive Processes (e.g., interpreting, comprehending, and applying) (e.g., Burstein et al., 2012; Keck, 2006; Mori, 2019; Shi, 2004; Shi & Dong, 2018; Wette, 2017).

The reliability and validity of the scale have been rigorously examined in Zhang (forthcoming) (see the footnote on p. 145). In the present study, confirmatory factor analysis (CFA) provided confirmatory evidence¹ that the first-order five-factor model was the best model ($\chi^2 = 552.684$, $df = 199$, $RMSEA = 0.077$ with 90% CI [0.070, 0.085], $CFI = 0.865$, $TLI = 0.844$, $SRMR = 0.074$) (see Figure 1). Cronbach's α coefficients among the 5 subscales ranged from .618 to .895, suggesting that the internal consistency of the scale was acceptable (Kline, 1999).

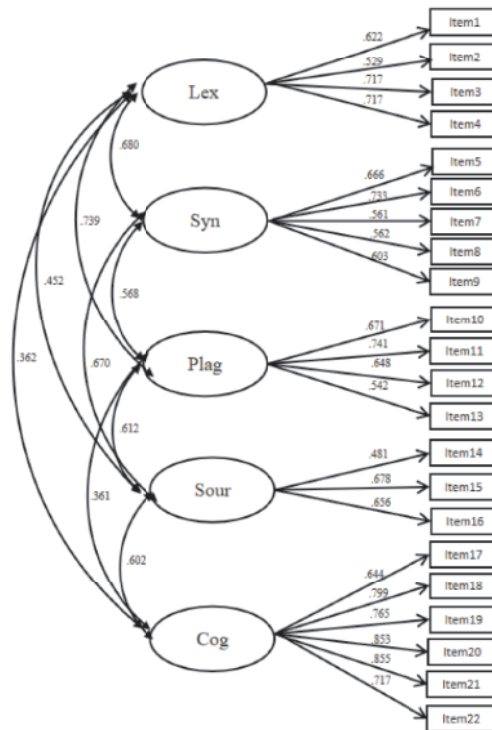


Figure 1. The structure of the first-order five-factor model.

3.2.2. Demographic variables

There are three demographic variables in the present study, including country, grade, and major. These factors were considered in analyses, as prior studies have indicated their associations with L2 paraphrasing competence (e.g., Keck, 2006, 2014; Shi, 2004; Shi & Dong, 2018; Wette, 2010; Zhang & Yu, 2023). Students were asked to report their countries (1 = China; 2 =

Japan), grade (1 = year 1 undergraduate students; 2 = year 2 undergraduate students; 3 = year 3 undergraduate students; 4 = year 4 undergraduate students; 5 = graduate students), and major (1 = hard discipline; 2 = soft discipline). To examine whether the latent profiles were associated with the demographic variables related to the participants, profile membership was specified as the predictor and these three demographic variables (country, grade, and major) as outcomes.

3.2.3. L2 proficiency

A self-report holistic scale was used to assess the participants' L2 proficiency, considering its advantage of being easy and fast to use (Weigle, 2002). The scale has a total of 6 items scored on a 5-point scale, which asked the participants to self-report their general L2 proficiency, L2 reading, L2 writing, L2 vocabulary size, L2 grammar knowledge and translating. The Cronbach's α is .710, showing that the reliability of the L2 proficiency scale was acceptable (Kline, 1999). To examine the association between profiles and L2 proficiency, profile membership was specified as the predictor and the six L2 proficiency measures as outcomes. The three above-mentioned demographic variables were included as covariates to control for their potential confounding influence, and to estimate the unique relationship between profile membership and L2 proficiency.

3.3. Data analytic approach

LPA was conducted on the 297 participants' self-reported responses in the survey to identify the groups of L2 learners with distinct profiles of paraphrasing competence patterns (Collins & Lanza, 2010). LPA is a clustering approach for identifying potentially homogeneous profiles. The participants' average factor scores were used as indicators to probabilistically assign students to their most likely latent profile. A series of fit indices was used to determine the optimal number of profiles. We used Akaike information criterion (AIC), Bayesian information criterion (BIC), sample-size-adjusted BIC (ABIC), Lo-Mendell-Rubin likelihood ratio test (LMRT), bootstrap likelihood ratio test (BLRT), and Entropy (Nylund et al., 2007; Peugh & Fan, 2013). Lower values of AIC, BIC, and ABIC were desirable in evaluating the model fit. Entropy indicates classification accuracy ranging from 0 to 1, with higher values indicating higher classification accuracy. The LMRT and BLRT compare the model fit of the K profile and K-1 profile.

If the p-value is significant, it indicates that the K profile provides a better model fit than the K-1 profile (Dziak et al., 2014). We also gave attention to the model interpretability, and a more parsimonious profile solution was preferred if an additional profile failed to provide more meanings (Muthen, 2004; Nagin, 2005). The LPA was performed using Mplus 8.4.

To examine the effects of identified profiles on country, grade, major, and L2 proficiency, the Bolck-Croon-Hagenaars (BCH) method was employed, which is recommended as this method profiles individuals based on posterior probabilities and adjusts the outcome analyses with classification errors being considered. It presents the effects of profile membership on outcomes in the form of pairwise differences between profiles in the means of the outcomes that are conditional on latent profile membership and demographic variables. Given that the BCH method integrates the most likely membership and classification errors, it has high statistical power and can precisely detect associations between profile membership and outcomes (Asparouhov & Muthen, 2014; Bakk & Vermunt, 2016).

4. Results

The results are presented in three parts. First, a latent profile analysis of five paraphrasing competence subscales identified three distinct learner profiles among the sample of Chinese and Japanese university students. Second, profile membership showed significant associations with the demographic variables of country and grade, but not major. Finally, after controlling for these variables, the three profiles were significantly associated with all six measures of L2 proficiency.

4.1. Profiles of Chinese and Japanese L2 learner paraphrasing competence

This subsection presents the profiles of Chinese and Japanese L2 learner paraphrasing competence derived from the LPA. Table 1 presents the LPA model fit indices. The AIC, BIC and ABIC values demonstrated a decreasing trend as the number of profiles increased. The LMRT values of the 4- and the 5-profile model were insignificant, suggesting that the 4-profile model is not significantly better than the 3-profile model and the 5-profile model is not significantly better than the 4-profile model. Although the 4-profile model showed the highest entropy among the five models, meaning that the

accuracy of this classification is the highest, eight participants out of the whole sample are too small for a single group (less than 5%). Therefore, the 3-profile model was preferred.

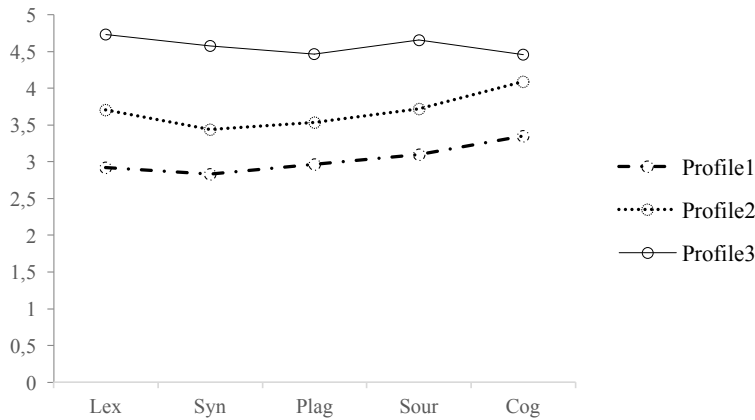
Number	Log likelihood	AIC	BIC	ABIC	Entropy	LMRT	BLRT	Class size per profile
Solutions for the five first first-order factors								
1	-1541.857	3103.715	3140.652	3108.939				297
2	-1447.481	2926.962	2986.061	2935.320	0.611	0.1429	0.0000	140/157
3	-1392.637	2829.275	2910.537	2840.768	0.801	0.0144	0.0000	81/196/20
4	-1338.274	2732.548	2835.973	2747.175	0.861	0.2534	0.0000	77/21/8/191
5	-1326.200	2720.401	2845.988	2738.162	0.805	0.8395	0.0128	5/68/57/146/21

Table 1. LPA model fit statistics.

Figure 2 illustrates three distinct profiles of L2 students’ paraphrasing competence. Specifically, Profile 1 (indicated by the black, dotted line) was characterized by the lowest levels across all five measured dimensions: Lexical Paraphrasing, Syntactic Paraphrasing, Plagiarism Avoidance, Source Use, and Cognitive Processes. This consistent pattern of low performance justified its label as “struggling paraphrasers” (SP), and it comprised 28% of the sample (N = 81).

Profile 2 (grey, dotted line) exhibited medium levels of Lexical and Syntactic Paraphrasing, Plagiarism Avoidance, and Source Use, but was distinguished by having the highest level of Cognitive Processes. This unique combination of average foundational skills with strong active processing led to its name, “average paraphrasers with active cognitive processes” (APACP). It constituted the majority of the sample at 65% (N = 196).

In contrast, Profile 3 (black, solid line) demonstrated the highest levels of Lexical Paraphrasing, Syntactic Paraphrasing, Plagiarism Avoidance, and Source Use. However, within this otherwise high-performing group, their level of Cognitive Processes was notably lower than that of Profile 2. This profile’s defining strength was its superior Lexical Paraphrasing score, which was the highest of all groups, leading to its label as “capable paraphrasers with strong lexical paraphrasing competence” (CPSLP). It represented 7% of the sample (N = 20).



Note. Lex. represents Lexical Paraphrasing; Syn. represents Syntactic Paraphrasing; Plag. represents Plagiarism Avoidance;

Sour. represents Source Use Strategies; Cog. represents Cognitive Processes.

Figure 2. Profiles of L2 students' paraphrasing competence (N = 297).

Subscales	SP		APACP		CPSLP	
	M	95%CI	M	95%CI	M	95%CI
Lex.	2.873	[2.760,2.987]	3.717	[3.644,3.790]	4.763	[4.534,4.991]
Syn.	2.795	[2.672,2.918]	3.447	[3.368,3.526]	4.580	[4.332,4.828]
Plag.	2.949	[2.838,3.060]	3.533	[3.462,3.604]	4.519	[4.296,4.742]
Sour.	3.049	[2.929,3.170]	3.730	[3.652,3.807]	4.700	[4.457,4.943]
Cog.	3.298	[3.178,3.419]	4.101	[4.024,4.179]	4.458	[4.215,4.701]

Note. * $p < .05$; ** $p < .01$; *** $p < .001$;

Note. SP represents struggling paraphrasers; APACP represents average paraphrasers with active cognitive processes;

CPSLP represents capable paraphrasers with strong lexical paraphrasing competence.

Table 2. Descriptive statistics of the 3 profiles on the 5 subscales of paraphrasing competence.

4.2. Associations between L2 learner paraphrasing competence profiles and demographic variables

To examine whether the identified paraphrasing profiles were associated with the participants' demographic variables, the BCH method was employed. As Table 3 shows, profile membership showed significant associations with the participants' country, with a small effect size, and with grade, with a small-to-medium effect size. No significant association was found with the participants' major.

	Means			Between-group mean differences					
	SP	APACP	CPSLP	2 versus 1		3 versus 1		3 versus 2	
				Est.diff	Cohen's d	Est.diff	Cohen's d	Est.diff	Cohen's d
Country	1.28	1.12	1.09	-.16**	-.36	-.19*	-.46	-.03	-.08
Grade	2.41	1.80	2.83	-.61***	-.44	.42	.28	1.03**	.77
Major	1.29	1.23	1.41	-.06	-.12	.12	.24	.18	.36

Note. **p* < .05, ***p* < .01, ****p* < .001.

Table 3. Means of country, grade, and major for each profile group and between-group mean differences (N = 297).

4.3. Associations between L2 learner paraphrasing competence profiles and L2 proficiency

Table 4 presents the results of the associations between L2 learner paraphrasing competence profiles and L2 proficiency after controlling for all three demographic variables. The results showed that the three profiles differed significantly on all six L2 proficiency measures. The effect sizes associated with profile differences (3 versus 2, 3 versus 1) in L2 writing, L2 vocabulary size and L2 grammar knowledge were large and medium in magnitude, respectively. The effect sizes associated with profile differences (3 versus 2, 2 versus 1) in general L2 proficiency were medium and small in magnitude, respectively.

	Means			Between-group mean differences					
	SP	APACP	CPSLP	2 versus 1		3 versus 1		3 versus 2	
				Est.diff	Cohen's d	Est.diff	Cohen's d	Est.diff	Cohen's d
L2	2.71	2.37	3.13	-.34**	-.37	.42	.43	.76**	.76
Reading	3.03	2.45	3.21	-.58***	-.62	.18	.19	.77**	.78
Writing	2.62	2.28	3.13	-.35**	-.40	.51*	.52	.86**	.85
Vocabulary	4.07	3.43	4.59	-.64***	-.71	.51**	.63	1.16***	1.30
Grammar	4.20	3.28	4.68	-.92***	-.99	.48**	.70	1.40***	1.60
Translating	4.12	3.45	4.15	-.68***	-.75	.03	.03	.71*	.65

Table 4. Means of L2 proficiency measures for each profile group and between-group mean differences (N = 297).

5. Discussion

5.1. Profiles of Chinese and Japanese L2 learner paraphrasing competence

In terms of the first research question concerning the profiles of Chinese and Japanese L2 learner paraphrasing competence, this study identified three profiles of L2 learner paraphrasing competence, that is, SP, APACP, and CPSLP. This suggests that most L2 learners are at the average level and have a large room for improvement.

Overall, all subscales in the CPSLP profile scored best, followed by the APACP profile, and then the SP profile. The largest variation shown among the three profiles is in the Lexical Paraphrasing subscale, followed by Syntactic Paraphrasing and Source Use, and the smallest variation shown among the three profiles is in Cognitive Processes, followed by Plagiarism Avoidance. In other words, L2 learners showed the largest differences in their reformulation of the original texts by making changes in words. SP and APACP demonstrated the highest competence on Cognitive Processes, while CPSLP showed the lowest competence on Cognitive Processes, suggesting that higher- and lower-level students differed in their cognitive processes or strategy use in paraphrasing. This distinction in cognitive processing is consistent with prior research, such as the finding by Ni, Chatupote, and Teo (2008) that successful and unsuccessful L2 learners employ different learning strategies. It is plausible that the low-level L2 students who had fewer language resources at their disposal had to resort to employing strategies more frequently to compensate for their limited knowledge (Yuan et al., 2004).

5.2. L2 learner paraphrasing competence profiles and demographic variables

Regarding the association between L2 learner paraphrasing competence profiles and the three demographic variables (country, grade, and gender), several key findings emerged. Firstly, students' country was statistically associated with L2 paraphrasing competence profile membership. However, this effect was nuanced: while the majority of both Chinese and Japanese participants were classified in the APACP profile, their proportional distribution across the other two profiles differed significantly. One possible reason is that the sample size had a relatively unbalanced country ratio between Chinese and Japanese participants (78% versus 22%), due to the difficulty of recruiting Japanese participants. This discrepancy may also stem from the differences in paraphrasing instruction between China and Japan. In China, the majority of college English teachers put emphasis on knowledge-telling paraphrasing and ask students to rephrase complex

sentences from textbooks in English class (Hirvela & Du, 2013). In Japan, while thesis writing has become a compulsory component in the 4-year Bachelor's programs at Japanese universities (Yamada, 2016), research found that Japanese university students received inadequate paraphrasing instruction (Yoshimura, 2015), because of the limited English academic writing support in Japanese universities (Rinnert & Kobayashi, 2005).

Secondly, grade was found to be significantly associated with L2 learners' paraphrasing competence, confirming the earlier findings which reported that undergraduate and graduate students exhibited different features in paraphrasing: undergraduate students were characterized by making little text revision (Keck, 2006, 2014; Shi, 2004), focusing on the knowledge-telling dimension of paraphrasing (Hirvela & Du, 2013) and the elevation of writing style instead of argumentation (Du & Liu, 2024), while graduate students could use paraphrasing to integrate what they read into their own arguments (Shi & Dong, 2018; Shi et al., 2018). The reason might be that students in higher grades might receive more academic writing instruction and gain more experience in academic writing, and therefore their paraphrasing competence might increase, echoing the instructional intervention effect in Wette (2010).

Thirdly, major had no significant association with profile membership. This is partly consistent with Shi et al. (2018), who also found no disciplinary difference in students' knowledge-telling paraphrases. However, the findings in the present study did not lend support to the significant effect of major on the students' knowledge-transforming paraphrases (Shi et al., 2018). The reason might be that, on the one hand, Shi et al. (2018) examined students' paraphrased texts and counted the frequencies of linguistic adaptations and content recontextualization, while the present study examined L2 students' self-reported paraphrasing competence, which mainly concerns their beliefs about what they can do instead of what they do in practice. It is possible that their beliefs and practices about paraphrasing mismatch. Hence, studies on the relationship between major and L2 learner paraphrasing competence are warranted, especially qualitative studies. On the other hand, the majority of the participants in this study were undergraduate students who had limited training in using paraphrasing as a source use device in academic writing, while participants in Shi et al. (2018) were all graduate students, who had received more training in paraphrasing than undergraduate students. Therefore, future studies could consider both undergraduate and graduate students with a more balanced sample size.

5.3. L2 learner paraphrasing competence profiles and L2 proficiency

As regards the association between profiles and L2 proficiency after controlling for the demographic variables, it was found that all L2 proficiency measures could significantly distinguish profiles, suggesting that L2 proficiency underlies paraphrasing competence. As learners' L2 proficiency improves, their linguistic processing capacity will improve as well, so that they can be more capable of paraphrasing. The finding meshes well with the prior studies that reported the relationship between paraphrasing competence and L2 proficiency (Basham & Rounds, 1984; Currie, 1998; Keck, 2006, 2014; Shi, 2004; Zhang & Yu, 2023). The finding in the present study indicates that concerning L2 learners whose L2 proficiency is still developing, research has shown that these learners may fail to understand the original texts or have no available resources to reformulate others' words (Hayes & Introna, 2005; Mori, 2019). Overwhelmed by the considerable distance between what is expected of them and what they can actually do (Hyland, 2001), low-proficiency L2 learners have no choice but to copy the chunks of original texts with no intention of plagiarism. Their heavy reliance on the source texts in their paraphrases might be "an important transitional strategy in the student's progress toward membership in a discourse community" (Howard, 1995, p. 788).

6. Conclusion

This study aimed to profile Chinese and Japanese L2 university students' paraphrasing competence and examine associations with key demographic and proficiency variables. The results are presented in three parts. First, a latent profile analysis of five paraphrasing competence subscales identified three distinct learner profiles among the participants: struggling paraphrasers, average paraphrasers with active cognitive processes, and capable paraphrasers with strong lexical paraphrasing competence. Second, profile membership showed significant associations with the demographic variables of country and grade, but not with major. Finally, after controlling for these variables, the three profiles differed significantly on all six measures of L2 proficiency.

This study yields important implications. English for Academic Purposes (EAP) researchers and teachers need to be cognizant of the complexity and multidimensionality of paraphrasing and avoid the dichotomous portrait of

either competent or incompetent paraphrase writers in judging L2 students' paraphrasing competence. More specifically, they need to pay attention to the development of all dimensions of paraphrasing competence, including Lexical Paraphrasing, Syntactic Paraphrasing, Plagiarism Avoidance, Source Use, and Cognitive Processes. For example, EAP teachers could emphasize the lexical and syntactic transformations in paraphrasing and combine vocabulary and grammar instruction with paraphrasing instruction in English academic writing courses to help students build declarative knowledge of paraphrasing. Strategy instruction, such as the techniques of paraphrasing, is also needed to help L2 learners translate the declarative knowledge of paraphrasing into the procedural knowledge of paraphrasing. Teachers may design different paraphrasing tasks, such as paraphrasing judgment, rewriting sentences and integrated writing to help students understand what paraphrasing is and practice how to paraphrase. English academic writing courses need to inform L2 learners (graduate students in particular) of plagiarism and source use issues involved in paraphrasing. Exemplars of paraphrasing texts from academic articles can be provided to students to analyze and model effective paraphrasing and learn to write legitimate paraphrasing in their own academic writing. Teachers may also use formative assessment to provide descriptive and diagnostic feedback to L2 learners on different dimensions of paraphrasing and encourage them to conduct self-assessment about their development of paraphrasing competence. These strategies have the potential to transform L2 learner paraphrasing competence from a low-level to a high-level profile.

It must be acknowledged that this study has several limitations. Firstly, the sample size is not large enough, and Chinese and Japanese subgroups were slightly imbalanced, calling for future studies to recruit more participants with a balanced sample size. Secondly, given that all constructs were measured with the participants' self-reported data, shared method and shared participant variance may inflate some of the identified associations, and therefore, multi-informant and multimethod designs are needed in future studies.

Acknowledgements

The authors wish to extend their sincere appreciation to the editor and the anonymous reviewers. Their thoughtful and constructive feedback was

invaluable in strengthening and refining this manuscript. We are also deeply grateful to the participants, whose willingness to share their time and perspectives made this study possible.

Article history:

Received 21 May 2025

Received in revised form 8 March 2026

Accepted 21 May 2026

References

- Asparouhov, T., & Muthen, B. (2014). *Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary Model* [Technical report comment]. Retrieved from https://www.statmodel.com/download/asparouhov_muthen_2014.pdf
- Bakk, Z., & Vermunt, J. K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling*, 23, 20-31. <https://doi.org/0.1080/10705511.2014.955104>
- Basham C., & Rounds, P. (1984). A discourse analysis approach to summary writing. *TESOL Quarterly*, 18(3), 527. <https://doi.org/10.2307/3586721>
- Burstein, J., Flor, M., Tetreault, J., Madhani, N., & Holtzman, S. (2012). *Examining linguistic characteristics of paraphrase in test-taker summaries*. ETS Research Report.
- 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 University Press.
- Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Wiley.
- Culicover, P. (1968). Paraphrase generation and information retrieval from stored text. *Mechanical Translation and Computational Linguistics*, 11, 78-88.
- Currie, P. (1998). Staying out of trouble: Apparent plagiarism and academic survival. *Journal of Second Language Writing*, 7(1), 1-18. [https://doi.org/10.1016/S1060-3743\(98\)90003-0](https://doi.org/10.1016/S1060-3743(98)90003-0)
- Du, Q., & Liu, Y. (2024). Foregrounding learner voice: Chinese undergraduate students' understanding of paraphrasing and source use conventions for English research paper writing. *Language Teaching Research*, 28(4), 1451-1471. <https://doi.org/10.1177/13621688211027032>
- Dziak, J. J., Lanza, S. T., & Tan, X. (2014). Effect size, statistical power and sample size requirements for the bootstrap likelihood ratio test in latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 534-552. <https://doi.org/10.1080/10705511.2014.919819>
- Hayes, N., & Introna, L. D. (2005). Cultural values, plagiarism, and fairness: When plagiarism gets in the way of learning. *Ethics and Behavior*, 15(3), 213-231. https://doi.org/10.1207/s15327019eb1503_2
- Hirvela, A., & Du, Q. (2013). "Why am I paraphrasing?": Undergraduate L2 writers' engagement with source-based academic writing and reading. *Journal of English for Academic Purposes*, 12(2), 87-98. <https://doi.org/10.1016/j.jeap.2012.11.005>
- Howard, R. M. (1995). Plagiarisms, authorships, and the academic death penalty. *College English*, 57, 788-806.
- Hyland, F. (2001). Dealing with plagiarism when giving feedback. *ELT Journal*, 55, 375-381. <https://doi.org/10.1093/elt/55.4.375>
- Keck, C. (2006). The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *Journal of Second Language Writing*, 15(4), 261-278. <https://doi.org/10.1016/j.jslw.2006.09.006>
- Keck, C. (2010). How do university students attempt to avoid plagiarism? *Writing & Pedagogy*, 2(2), 193-222. <https://doi.org/10.1558/wap.v2i2.193>
- Keck, C. (2014). Copying, paraphrasing, and academic writing development: A re-examination of L1 and L2 summarization practices. *Journal of Second Language Writing*, 25, 4-22. <https://doi.org/10.1016/j.jslw.2014.05.005>

- Kline, P. (1999). *The handbook of psychological testing*. Routledge.
- Martin, R. (1976). *Inférence, antonymie et paraphrase*. Klincksieck.
- Mel'čuk, I. (1992). Paraphrase et lexique: La théorie sens-texte et le dictionnaire explicatif et combinatoire. In I. Mel'čuk, N. Arbatchewsky-Jumarie, L. Iordanskaja, & S. Mantha (Eds.), *Dictionnaire explicatif et combinatoire du Français contemporain. Recherches lexico-sémantiques III* (pp. 9-58). Presses de l'Université de Montréal.
- Moore, T. (1997). From text to note: Cultural variation in summarization practices. *Prospect*, 12, 54-63.
- Mori, M. (2019). Pluralizing the paraphrase: Reconsidering meaning via a typology of linguistic changes. *TESOL Quarterly*, 53(3), 885-895. <https://doi.org/10.1002/tesq.514>
- Muthén, B. (2004). *Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data*. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 345-370). Sage.
- Nagin, D. S. (2005). *Group-based modeling of development*. Harvard University Press.
- Ni, Q., Chatupote, M., & Teo, A. (2008). A deep look into learning strategy use by successful and unsuccessful students in the Chinese L2 learning context. *RELJ Journal*, 39, 338-358. <https://doi.org/10.1177/0033688208096845>
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 535-569. <https://doi.org/10.1080/10705510701575396>
- Peugh, J., & Fan, X. (2013). Modeling unobserved heterogeneity using latent profile analysis: A Monte Carlo simulation. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(4), 616-639. <https://doi.org/10.1080/10705511.2013.824780>
- Rinnert, C., & Kobayashi, H. (2005). Borrowing words and ideas: Insights from Japanese L1 writers. *Journal of Asian Pacific Communication*, 15(1), 31-56. <https://doi.org/10.1075/japc.15.1.05rin>
- Shi, L. (2004). Textual borrowing in second-language writing. *Written Communication*, 21(2), 171-200. <https://doi.org/10.1177/0741088303262846>
- Shi, L., & Dong, Y. N. (2018). Chinese graduate students paraphrasing in English and Chinese contexts. *Journal of English for Academic Purposes*, 34, 46-56. <https://doi.org/10.1016/j.jeap.2018.03.002>
- Shi, L., Fazel, I., & Kowkabi, N. (2018). Paraphrasing to transform knowledge in advanced graduate student writing. *English for Specific Purposes*, 51, 31-44. <https://doi.org/10.1016/j.esp.2018.03.001>
- Sun, Y. C., & Yang, F. Y. (2015). Uncovering published authors' text-borrowing practices: Paraphrasing strategies, sources, and self-plagiarism. *Journal of English for Academic Purposes*, 20, 224-236. <https://doi.org/10.1016/j.jeap.2015.05.003>
- Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.
- Wette, R. (2010). Evaluating student learning in a university-level EAP unit on writing using sources. *Journal of Second Language Writing*, 19, 158-177. <https://doi.org/10.1016/j.jslw.2010.06.002>
- Wette, R. (2017). Source text use by undergraduate post-novice L2 writers in disciplinary assignments: Progress and ongoing challenges. *Journal of Second Language Writing*, 37, 46-58. <https://doi.org/10.1016/j.jslw.2017.05.015>
- Xu, J., & Zheng, Y. (2025). Negotiating understanding, control, and authorship: L2 learners' experiences with AI-assisted paraphrasing in academic writing. *Journal of English for Academic Purposes*, 78, 101591. <https://doi.org/10.1016/j.jeap.2025.101591>
- Yamada, K. (2003). What prevents ESL/EFL writers from avoiding plagiarism? Analyses of 10 North American college websites. *System*, 31, 247-258. [https://doi.org/10.1016/S0346-251X\(03\)00023-X](https://doi.org/10.1016/S0346-251X(03)00023-X)
- Yamada, K. (2016). Cross-disciplinary variations: Japanese novice writers' socialization into the undergraduate thesis. *The Asia-Pacific Education Researcher*, 25, 207-217. <https://doi.org/10.1007/s40299-015-0252-3>
- Yoshimura, F. (2015). Japanese university students' task representations of paraphrasing and their experience with it. *Tohoku Gakuin University Review: Essays and Studies in English Language and Literature*, 99, 1-21.
- Yuan, F., Liu, Z., & Zhang, F. (2004). A study of the differences in strategy use between English majors and non-English majors. *Foreign Language World*, 104, 25-32.
- Zhang, E. D. (2020). An investigation of novice

ESL writers' cognitive processes and strategy use of paraphrasing. *Language Testing in Asia*, 10(9), 1-26. <https://doi.org/10.1186/s40468-020-00104-3>

Zhang, E. D., & Yu, S. (2023). Investigating the relationship between linguistic changes in L2

writers' paraphrasing, paraphrasing performance and L2 proficiency. *Applied Linguistics Review*, 14(5), 1451-1473. <https://doi.org/10.1515/applirev-2021-0014>

Emily Di Zhang is an Assistant Professor at the School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China. Her research interests include L2 writing assessment. She has publications in journals like *Computer Assisted Language Learning*, *Journal of Second Language Writing*, *System*, *Language Teaching*, *Assessing Writing*, and *Language Teaching Research*.

Nan Zhou is an Associate Professor at Faculty of Education, University of Macau, Macau SAR, China. His research focuses on child and youth development and particular adjustment problems in family and peer contexts. His publications have appeared in *Journal of Child Psychology and Psychiatry*, *Developmental Psychology*, *Journal of Research on Adolescence*, etc.

Shulin Yu is an Associate Professor at Faculty of Education, University of Macau, Macau SAR, China. His research interests include second language writing and classroom feedback and assessment. His publications have appeared in *Educational Research Review*, *Assessing Writing*, *Journal of Second Language Writing*, *Language Teaching Research*, *Language Teaching*, and *TESOL Quarterly*.

NOTES

¹ The forthcoming paper deals with the development and validation of an L2 student paraphrasing competence scale. We conducted the exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) to validate the scale. The final version of the scale comprises 22 items and five dimensions, i.e., Lexical Paraphrasing, Syntactic Paraphrasing, Plagiarism Avoidance, Source Use, and Cognitive Processes (see Appendix III).

Appendix

I. Please provide the information below.

1. Country _____
2. Major _____
3. Grade _____

II. Please self-assess your English L2 proficiency based on the five-point holistic scale.

1. Please rate your general English proficiency.
2. Please rate your English reading proficiency.
3. Please rate your English writing proficiency.
4. Please rate your English vocabulary knowledge.
5. Please rate your English grammar knowledge.
6. Please rate your English translating proficiency.

III. Please self-report your English L2 paraphrasing competence in English based on the five-point Likert scale.

1. I can use synonyms to substitute the words in the source texts.
e.g., She is beautiful → She is pretty.
2. I can use antonyms to substitute the words in the source texts. e.g., My health is good → My health is not bad.
3. I can change the morphology of the words in the source texts.
e.g., I was surprised at what he did. → What he did was a surprise to me.
4. I can convert one word in the source texts into a multiple-word unit, or vice versa. e.g., I came up with a good idea. → I had a great idea.
5. I can convert the active voice in the source texts into the passive voice.
e.g., I made the video. → The video is made by me.
6. I can use conjunctions like and/but/so to turn complex sentences in the source texts into compound sentences or merging simple sentences in the source texts.
7. I can downgrade main clauses in the source texts to dependent clauses, e.g., it is true → although it is true.
8. I can convert relative clauses in the source texts into phrases, and vice versa, e.g., directions that might not work → the wrong directions; to dominate discussions → who consistently control conversations.
9. I can convert noun phrases in the source texts into pronouns.
e.g., This room is bright. → It is bright.
10. I understand that paraphrasing is a source use strategy to avoid plagiarism.
11. I can distinguish between legitimate paraphrases and superficial patchwriting that risks plagiarism.
12. I can substantially transform the original texts to avoid copying.

13. I can give appropriate citations to the source texts.
14. I can omit irrelevant original information to support my original argument.
15. I can interpret the source texts to recontextualize the paraphrases.
16. I can add my own ideas to the paraphrases for rhetorical purposes.
17. I need to comprehend the meaning of the source texts before writing paraphrases.
18. I need to analyze the source texts in paraphrasing, such as the importance of ideas and logic of the source texts.
19. I need to synthesize the source texts in paraphrasing.
20. I need to apply my linguistic knowledge in writing paraphrases.
21. I need to constantly check the quality of paraphrases, such as the transformation of source language, fidelity to source meaning, linguistic accuracy, and its rhetorical effectiveness.
22. I need to evaluate the quality of the final paraphrases, such as the transformation of source language, fidelity to source meaning, linguistic accuracy, and its rhetorical effectiveness.

