

The Emergence of a Complex Language Skill: Evidence from the Self-organization of Interpreting Competence in Interpreting Students

Research Article

Zhibin Yu¹ and Yanping Dong² 

Cite this article: Yu Z, Dong Y (2022). The Emergence of a Complex Language Skill: Evidence from the Self-organization of Interpreting Competence in Interpreting Students. *Bilingualism: Language and Cognition* **25**, 269–282. <https://doi.org/10.1017/S1366728921000870>

Received: 4 March 2021
Revised: 6 August 2021
Accepted: 27 August 2021
First published online: 28 September 2021

Keywords:

interpreting competence; self-organization; structural equation modeling; language competence; working memory; Dynamic Systems Theory

Author for correspondence:

Yanping Dong, E-mail: ydpdong@zju.edu.cn

¹Renji Faculty, Wenzhou Medical University, Wenzhou, China and ²Department of Linguistics, Zhejiang University, Hangzhou, China

Abstract

Research on the development of interpreting competence could be a window to the issue of how L2 learners develop complex language skills. The present study conducted a longitudinal experiment with beginning interpreting students, exploring the change of relationship between consecutive interpreting (CI) competence and two related capacities (i.e., language competence and memory capacity). Two major results were revealed. First, in general, more language skills and working memory (WM) spans got correlated with CI performance at the later stage of CI training. Second, a fit structural equation model of CI competence could only be reported in the post-test. We may therefore conclude that the development of interpreting competence is at least partly a result of the self-organization of the interpreting competence system, in which relevant components get mobilized, and a better coordinated structure emerges. Implications for the development of complex language skills and for the concept of self-organization are discussed.

1. Introduction

The development of language competence is an intricate process involving various cognitive factors, and the issue of how language learners acquire complex language skills during this intricate process has been one of the central topics in language acquisition research. Given that language is a complex system whose performance depends on how language users “soft-assemble” available resources (Larsen-Freeman, 2018), to unveil how complex language skills are developed, researchers are essentially required to probe into the interaction of related cognitive resources, as well as its change over time. Among the available language acquisition theories, we consider the DYNAMIC SYSTEMS THEORY (DST) a promising META-THEORY particularly suitable for addressing this issue, as it recognizes the role of interactions of variables as the crucial intrinsic driving force for development (de Bot, Lowie & Verspoor, 2007; Larsen-Freeman, 1997; van Geert, 2009).

Being one of the most complex language skills, interpreting could be a window to the mystery of language competence development. To satisfactorily fulfill the interpreting task demands, interpreters (in particular, interpreting students) are required to mobilize all their relevant cognitive abilities (Dong, 2018; see Göpferich, 2013 for a similar proposal for translators). However, little research up to date has explored how these different cognitive abilities are mobilized in interpreting training, as well as how these abilities interact to produce the intended progress. The present study, therefore, intends to address this issue in the framework of the DST, hoping to reveal how interpreting competence develops in interpreting students as the interrelationship among involved cognitive abilities evolves within the complex system. This exploration may contribute to research on human development by shedding light on how complex language skills emerge in language learners.

1.1. The DST approach in the study of language skill development

The DST, as a meta-theory adopted in various fields of science, essentially concerns the major feature of a complex dynamic system, i.e., its change over time (de Bot et al., 2007; Dong, 2018; van Geert, 2008). Unlike traditional developmental theories (e.g., information processing model) which assume a predictable and somewhat linear path of development with a clear beginning and end state for each individual, the DST views the evolution of complex systems (such as human development) as a non-linear process with no clear end state. Individual variation, instead of being treated as noise, is deemed natural and important, resulting from the continuous coupling between the system and its environment (e.g., Larsen-Freeman, 2018; van Geert, 2009). For the study of complex systems, the DST highlights the important role

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

of interactions among sub-components in the emergence of a relatively stable state (an attractor state in its term), and considered it inadequate to investigate the development of the whole system by only examining its parts piecemeal (Larsen-Freeman, 1997).

Given its advantage over traditional approaches in accounting for non-linear behaviors, unpredictable outcomes and even messy facts, the DST has been recently applied to language acquisition research to study the development of language skills in L2 learners. These studies, most of which were targeted at L2 writing (e.g., Larsen-Freeman, 2006; Verspoor, Schmid & Xu, 2012), but also a few at L2 reading (Wang, 2011), L2 listening (Dong, 2016) and L2 speaking (e.g., Larsen-Freeman, 2006; Yu & Lowie, 2019), have probed into individual variation, phase transition and non-linearity manifested during the developmental process of different language skills. By analyzing the linguistic outputs (e.g., sentence length and grammatical features in speaking and writing) or processing strategies (e.g., summarization and note-taking in reading and listening) at a number of time points during L2 learning, these studies converged on the findings that language skill development is a dynamic, non-linear process with both progress and attrition, characterized by self-adaptation and self-restructuring, and sensitive to learners' initial state.

Despite the revealing findings yielded in the literature, one critical issue remains relatively unclear, i.e., how relevant cognitive abilities INTERACT in the complex dynamic systems of language learners, giving rise to the emergence of language skills. Language performance, according to the DST, depends on all the linguistic resources that learners could effectively assemble under their cognitive constraints, thus language competence should be viewed as a complex system consisting of not only its linguistic components (i.e., sub-systems like phonemes, morphemes, lexicon, syntax, etc.) but also relevant cognitive resources (e.g., memory capacity) that support (or constrain) language performance. Previous studies adopting the DST perspective have explored the evolution of sub-systems, as well as their interactions, at the linguistic level, but little research has probed into the change of inter-componential relationship at the cognitive level. The present research intends to contribute to the issue of language skills development from this perspective.

1.2 Interpreting as a complex language skill and research on its development

Interpreting is considered one of the most challenging language tasks (e.g., Christoffels & de Groot, 2005). Interpreters generally need to utilize all relevant cognitive abilities, such as language skills (including both comprehension and production), memory skills and executive functions, to fulfill the task requirements satisfactorily (Dong, 2018; Gile, 2009). This makes interpreting a good target for probing into the secrets of how a complex language skill emerges in training, resulting from the mobilization and interaction of its componential cognitive abilities.

Although research on interpreting is numerous, studies addressing the development of interpreting competence using empirical methods are limited. Some of these studies (e.g., Riccardi, 2005) adopted the expert-novice paradigm to identify the sub-competencies/sub-skills that were enhanced in interpreting training by comparing the application of interpreting strategies and/or interpreting performances between beginners and professional interpreters. The general conclusion was that the development of interpreting competence is a process of

automatizing strategy implementation, with a shift of primary dependence from knowledge-based strategies to skill-based strategies. Another approach explored the order in which different sub-competencies/sub-skills are assumed to be acquired or developed. These studies, either based on interpreting teaching practice (e.g., Liu, 2017) or in-depth interviews with professional conference interpreters (Albl-Mikasa, 2013), converged on the following order for the three components of interpreting competence: language competence, interpreting-specific skills and business competence. Specifically, language competence is a fundamental precondition for interpreting training, while interpreting-specific skills are mainly learned in training programs and further cultivated through work experiences. As for business competence, it is mainly acquired in the professional on-the-job phase. Almost all relevant previous research was targeted at the improvement of certain sub-skills (e.g., interpreting strategies). Given that interpreting is such a complex task that involves the interaction and coordination of many cognitive abilities, it seems necessary to explore the evolving inter-relationship among interpreting competence and relevant cognitive abilities, an approach suggested by the DST.

The idea of adopting the DST approach to investigate the development of interpreting (or translation in a broader sense) competence has been proposed in a few recent theoretical studies. For instance, Dong (2018) proposed that it might be revealing to probe into the SELF-ORGANIZATION of trainees' interpreting competence system by investigating the evolving interrelationship among its relevant cognitive components. Self-organization, a key concept in the DST, refers to a process in which some form of an overall pattern or coordination arises out of local interactions between component parts of an initially disordered system (van Geert, 2008; van Geert & Verspoor, 2015). In other words, it is the spontaneous occurrence of structured patterns, a process in which a system becomes more organized and coordinated through interactions of its components. This concept dates back to the belief of "spontaneous order" held by ancient Greek atomists, and was introduced into contemporary science by Ashby (1947). It became a fundamental notion in general systems theory (Von Bertalanffy, 1993), and has been adopted in a wide range of fields, including physics (e.g., self-assembly of nanoparticles), chemistry (e.g., molecular self-assembly), biology (e.g., flocking behavior in birds and fish) and cognition (e.g., neural organization). Recently, this concept has been applied in connectionism, a new approach targeting the learning and development in brains and brain-like computers (Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett, 1996), and recognized as one of the most important emergentist mechanisms (MacWhinney & O'Grady, 2015).

Although the working definitions may not be the same, three essential ideas concerning development can be identified in the notion of "self-organization" in various fields of research. First, development is a process of change, which entails differences being observed in a dynamic system between different stages. The change not only occurs in individual components, but also in the connections and interactions among components (Kelso, 1995). Second, the system evolves from an initially chaotic state to a relatively ordered one, with a pattern of inter-componential relations emerging during this process. Such a change could be indicated either by the formation of connections among originally separate components, or by the optimization of the system's organization from a bad to a good one (Von Bertalanffy, 1993). Third, the formation of pattern/order/connections and the realization of coordination among components are the spontaneous results of coupling between the system and the environment,

not something prescribed by certain internal parameters or external agents at the very beginning.

The notion of self-organization, therefore, could be an insightful perspective for the issue of how a complex language skill like interpreting emerges in interpreting trainees. However, little empirical research has been conducted from this perspective up to date. Consequently, despite relevant theoretical elucidations in the literature, it remains unexplored HOW SUCH A SELF-ORGANIZING PROCESS MAY BE MANIFESTED in the development of interpreting competence.

To be specific, two issues await to be investigated. First, what cognitive sub-competences or sub-skills are mobilized in different phases of interpreting training? According to the DST, self-organizing systems may incorporate more elements and maintain a more sophisticated arrangement of components as time goes on (Lewis, 2000; van Geert & Verspoor, 2015). This implies that different sub-competences may become contributive to interpreting performance at different stages, but the specific component(s) recruited in interpreting tasks at a given stage can only be revealed by empirical studies. Göpferich (2013) was probably the first and the only empirical research that addressed the mobilization of cognitive components in the framework of the DST. This study investigated the development of students' written translation competence in a 3-year program by analyzing their think-aloud transcriptions and questionnaire answers, and compared their translation products and processes with those of professional translators. The results indicated that different components were incorporated in the complex systems of translation competence in novices and experts. What was crucial for professional translators were three types of translation-specific sub-competence: strategic competence, translation routine activation competence, and tools and research competence. For students of translation, bilingual competence was much more important, whereas the three translation-specific sub-competences remained undeveloped throughout the three years of training. However, as the author mentioned, the variables analyzed were too few to reveal a comprehensive picture. More importantly, the cognitive components recruited in (oral) interpreting and (written) translation are probably not the same, and further research on interpreting is thus warranted.

Second, what evidence can be provided to illustrate a state of better connection and cooperation achieved among recruited cognitive components in an initially disordered system? Self-organization, by its definition, denotes the spontaneous emergence of patterns in a dynamic system. Given that the patterns yielded may vary across different systems depending on the systems' nature and scale, researchers in different fields seek for different types of evidence to describe and explain self-organization. For instance, while synchronization between metronome and limb was taken as the critical evidence for the action-perception coordination, coherent relations between neuronal activities were used to support self-organizing dynamics of the nervous system (Kelso, 1995). When it comes to language learning, previous studies tend to probe into the distinctive features of linguistic output manifested at different stages. For instance, Li, Zhao, and MacWhinney (2007) simulated a variety of lexical comprehension and production patterns in children's acquisition of words based on the self-organizing map (SOM) algorithm of Kohonen (2001). Verspoor, Lowie, and van Dijk (2008) probed into the amount and type of intra-individual variability in average word and sentence length to depict the developmental trajectory of L2 writing skills. Findings from these studies help reveal how

input characteristics interact with individuals to impact on the self-organizing process of learners' language competence systems. However, research on the change of linguistic output alone may not be sufficient to reveal the full picture of how learners develop a complex language skill, as the successful completion of a demanding language task like interpreting requires much more than mere proficient working languages. Thus studies probing into how efficient cooperation is achieved between relevant cognitive components, such as listening skills and working memory (WM) spans, are definitely in need for a more comprehensive view.

1.3 The present study

In the framework of the DST, the present study aims to investigate how interpreting competence develops towards interpreting training by addressing the above two issues. In accordance with the DST, interpreting competence is defined as a multi-componential adaptive system, in which all its sub-competences interact and cooperate so as to fulfill processing demands in interpreting tasks. Two hypotheses were generated based on the DST's account of development, one for each unresolved issue respectively. According to the DST, "[development] is a directed process, from an immature to a mature state, implying increasing complexity in terms of a system that incorporates more and more elements and at the same time integrates them" (van Geert & Verspoor, 2015, p. 537). This statement suggests two crucial changes that characterize development: first, a system becomes more complex by incorporating and integrating more elements; second, the system becomes more stable and mature, probably due to the better coordination achieved among internal elements. In line with this account, for the first issue, we hypothesized that more cognitive abilities would be recruited in the interpreting competence system after training, indicated by an increase of cognitive abilities that correlate with interpreting performance. For the second issue, we hypothesized that no fit model of interpreting competence with recruited cognitive abilities as its components can be established at the very beginning, while a fit model can be yielded after interpreting training. These two hypotheses were going to be tested in the present study by adopting empirical methods with a longitudinal design.

Albl-Mikasa (2013) suggests that formal or systematic parts of competence development are probably confined to the early stages. We therefore focused on interpreting trainees in their first year of consecutive interpreting (CI) training based on the following two concerns. First, these first-year trainees had just started their transition from general bilinguals to interpreters, and drastic reconstruction was likely to take place in their interpreting competence system. Second, since CI training is what interpreting trainees mainly (if not exclusively) receive at the beginning stage in most interpreting training programs, and is often considered a basis for simultaneous interpreting training, CI training was considered in the present study the appropriate mode for research on the initial development of interpreting competence.

Being multi-componential in nature, the interpreting competence system involves various cognitive abilities. The present research intends to focus on the dynamic roles of two fundamental components, i.e., language competence and working memory, and is going to address the issue of self-organization by exploring and comparing how these two sub-competences interact and contribute to CI performance at the beginning (Stage 1) and end

(Stage 2) of the academic year (i.e., respectively 2nd and 10th month of training). The significance of these two components in interpreting has been highlighted in theoretical models (e.g., Gile, 2009; PACTE, 2011) and recognized in empirical studies (e.g., Cai, Dong, Zhao & Lin, 2015; Christoffels, de Groot & Kroll, 2006; Tzou, Eslami, Chen & Vaid, 2012). For instance, Cai et al. (2015) reported that L2 proficiency measured before interpreting training (at least partly) predicted CI performance score a year later. However, no previous research has explored how these two fundamental components interact and work on interpreting performance, and how the interactive pattern gets optimized, giving rise to an improvement of interpreting competence.

Language proficiency and WM are both multi-facet constructs with many sub-skills or sub-components, and these sub-components may have different relationship with interpreting competence at different phases. For example, Dong, Cai, Zhao, and Lin (2013) revealed that only part of the measured subskills (e.g., SL comprehension, L2 listening span) correlated with CI performance at the 10th month of CI training. Were these subskills also involved at the very beginning of training, or only became mobilized after training? Would a different interactive pattern appear after training among these involved sub-skills? Answers to these questions would contribute to our understanding of the dynamic relationship between interpreting competence and its two fundamental components: language competence and working memory.

To evaluate CI competence, the present study assessed students' performance of both interpreting directions (E-C: from L2 English to L1 Chinese; C-E: from L1 Chinese to L2 English). According to the curriculum of university interpreting courses available to us, more CI training is arranged in the E-C rather than C-E direction, as recommended by experts in conference interpretation training (Seleskovitch, 1999). The difference in interpreting direction may show up not only in interpreting competence but also in its development.

To sum up, the present study aimed to explore how interpreting trainees' language and memory skills would get closely involved in CI tasks at the 2nd (Stage 1) and 10th month (Stage 2) of training, and to test whether a fit CI competence model could be yielded at these two stages. Based on the DST's account of development, we predicted that more language skills and/or working memory spans would become correlated with CI performance at Stage 2 than at Stage 1, and for E-C direction than for C-E direction (Prediction 1), and a fit model of CI competence could only be reported at Stage 2 (Prediction 2).

2. Materials and methods

2.1 Participants

Sixty-nine third-year undergraduate students (12 males) from a key university of foreign studies in China participated, and were paid after the experiment. Eight of them either quit the experiment or did not complete the task sets, thus statistical analyses were conducted on the data from the remaining 61 participants (10 males). These participants were all English majors specializing in interpreting and translation, with an English learning history of about 10 years. In accordance with the syllabus, they were in their first year of interpreting training, in which they needed to take 4 courses in interpreting and 4 courses in translation, together with other English courses such as literature. Each course lasted for 18

weeks, with 80-minute class time per week. All the participants had normal or corrected-to-normal vision and reported no hearing deficits or language disorder problems.

2.2 Tasks, procedure and scoring

To reveal the complex nature of interpreting competence, the present study adopted a variety of measures for language competence and working memory. For language competence, general L2 proficiency was measured, as well as three language subskills that were considered relevant to interpreting in the literature, i.e., lexical retrieval efficiency (Christoffels, de Groot & Waldorp, 2003), source language comprehension (Mayor, 2015) and important information selection in summary writing (Liu, Schallert & Carroll, 2004). For working memory, altogether seven WM span tasks were administered, including listening, reading and speaking spans in both L1 Chinese and L2 English, and a digit span task. This is due to two reasons. First, previous studies indicated that both encoding modality (reading, listening or speaking) and encoding language (L1 or L2) had an effect on WM spans (e.g., Ikeno, 2006; Lehnert & Zimmer, 2008). Second, Dong et al. (2013) and Cai et al. (2015) showed that WM spans measured by different tasks showed distinct relationship with CI performance. Altogether 13 tasks were administered, measuring participants' language skills, memory capacity and CI performance, both at Stage 1 and Stage 2 following the same procedure in the same order.

Tasks examining CI performance

Two validated CI tasks, each taking about 25 minutes, were administered, measuring participants' E-C and C-E CI performances respectively. Participants were required to interpret a conference speech excerpt from English into Chinese, or vice versa. Each speech took about 8 minutes, and both were segmented into appropriate length¹. The English speech was delivered by a native English speaker at an average rate of 143 words per minute, and the Chinese speech was delivered by a native Chinese speaker at an average rate of 264 characters per minute. During the tasks, participants listened to each segment of the speech one at a time, and began to interpret when they heard a sound signal indicating the end of each segment. The time set for interpretation after each segment was 1.5 times longer than the segment duration. Participants were allowed to take notes and refer back to these notes in their interpreting. The speech was presented aurally over headphones, and all participants' interpreting products were recorded on the computer. Participants' CI performances were rated by two professional interpreters with years of experience in both interpreting teaching and practice. Analytic rating scale was adopted for assessing the CI quality, with information accuracy and completeness accounting for 67% of the final score and TL grammar and appropriateness taking up 33%. Details of the test (i.e., materials, procedure and scoring) were reported in Cai

¹There were altogether 21 segments in the E-C CI materials, with each segment, as a complete sense unit, consisting of about 56 words on average. As for the C-E CI material, 20 segments were created, each consisting of 102 characters on average. Compared with common CI training practice in Europe, the segments in these two CI tasks were short. But given that our trainees were unbalanced bilinguals who had just started CI training, longer segments would be too demanding for them. The difficulty of the two CI tasks (in such segmentation) was considered appropriate based on evidence collected from a pilot study, judgments from five experienced interpreting instructors, and a questionnaire on the appropriateness of materials administered after the test (Dong et al., 2019).

et al. (2015) and Dong, Li, and Zhao (2019). The inter-rater reliability was excellent ($r = .89$).

Tasks testing language competence

Language competence was evaluated by five tests. The first two (TEM 4 and TEM 8) were typical comprehensive language (English) proficiency tests. The third, fourth and fifth tests respectively evaluated one's listening comprehension, summary writing and word access.

TEM 4 & TEM 8

Participants' scores on the Test for English Major Band Four (TEM4) and Band Eight (TEM 8) were respectively taken as indicators of their L2 proficiency around Stage 1 and Stage 2. These two criterion-referenced tests, administered nationally once a year in China, measured English majors' general English proficiency at the intermediate and advanced levels. As one of the few predominant large-scale English tests in China, TEM 4 and TEM 8 are considered valid and reliable (for a detailed description of these tests, see Jin & Fan, 2011).

SL listening comprehension test

Two short listening comprehension tests (one for E-C and one for C-E) were designed to examine participants' SL listening comprehension. There were eight test items, generated by the designers of the two CI tests according to the CI task materials (see Dong et al., 2019 for introduction). Each item was assigned one point in scoring.

Summary writing for SL

A summary writing task was developed by the designers of the CI tests (see Dong et al., 2019 for introduction) to examine participants' SL summarization ability. Participants were asked to write two short summaries of the speech they had just interpreted in the CI tests, one for each direction within 150–200 words. The summary writing task, taking 15 minutes, was administered immediately after the completion of each CI task. As for the scoring criterion, two interpreting teachers were asked to select the key information in the speech, and the final version of the suggested answers included nine information units that both teachers viewed as key points. Each of the nine units contained two sub-units, with each sub-unit assigned 0.5 point in scoring.

Word translation recognition task

A computer-based word translation recognition task in both L1-L2 and L2-L1 directions, programmed by E-Prime 2.0, was adopted to measure participants' lexical retrieval efficiency. Two lists of word pairs, each with 100 Chinese and 100 English words, were compiled as the stimuli (see Cai et al., 2015 for more details) and used in the C-E and E-C sessions respectively. In each list, half of the word pairs were translation equivalents. The order of the two sessions was randomized across participants to avoid a potential order effect of the translation directions. Each trial in the task started with a fixation ('+'), appearing at the center of the screen for one second, and then a translation word pair was presented. In the C-E session, a Chinese word was first presented slightly above the center of the screen. Fifty milliseconds later, an English word appeared below it. In the E-C session, the sequence was reversed. Participants were required to judge as rapidly and accurately as possible whether these two words were translation equivalents or not by pressing either the "F" or

the "J" key. Participants' judgment latency (i.e., response time, or RT) and accuracy were recorded.

Tasks measuring working memory

As justified above, seven computer-based memory tasks were adopted, programmed by E-Prime 2.0, targeting different dimensions of memory capacity.

English & Chinese reading span

The two span tasks differed only in the stimuli presented: English or Chinese sentences. In the English span task (Unsworth, Heitz, Schrock & Engle, 2005), participants were required to read sets of English sentences (presented one by one on the computer screen), make correctness judgment for each sentence, and remember a random letter presented after the judgment. At the end of each set (with the set sizes ranging from 3 to 7 sentences), participants were presented with a 4×3 matrix of letters (H, J, N, P, R, S, Q, F, K, L, T, Y) on the screen, and they were instructed to recall the letters by clicking the box next to the appropriate letters in the exact order of presentation. The number of correctly recalled letter (i.e., correct both in letter and in order) was taken as the reading span (maximum score = 75). As the original task was targeted at English natives, we conducted a norming test before the test proper on a new group of 28 participants from the same population, and replaced sentences that might be too difficult for our unbalanced bilingual participants (with average accuracy below 90%).

English & Chinese listening span

The English listening span tasks were an adapted listening variation of the task first developed by Daneman and Carpenter (1980). Participants in this task were required to listen to sets of sentences through earphones, remember the last word of each sentence and make correctness judgement. At the end of each set (with the set sizes ranging from 2 to 6 sentences), participants were asked to recall the words they remembered. The listening span score was calculated as the cumulative number of words correctly recalled from all sets (maximum score = 60). A norming test was administered beforehand on a new group of 26 participants to ensure the appropriateness of sentence difficulty (with average accuracy above 90%). The Chinese reading span task was almost the same as the English version, except that Chinese sentences were constructed and used as the materials.

English & Chinese speaking span

The English speaking span task adopted the paradigm first developed by Daneman and Green (1986). The stimuli were 100 English words containing two syllables and seven to eight letters. To ensure participants' familiarity with these words, a norming test was conducted beforehand on a new group of 20 students from the same population, and difficult words were replaced. In this task, sets of English words, presented one at a time, were shown in the center of a computer screen for participants to remember. At the end of each set (with the set sizes ranging from two to six words), a visual signal was presented telling participants to orally generate sentences containing these words. Participants' English speaking span was calculated as the total number of correctly recalled words in acceptably generated sentences (maximum score = 100). In the Chinese speaking span task, the materials were two-character Chinese words of high

frequency, taken from Modern Chinese Frequency Dictionary (Beijing Language Institute, 1986).

Digit Span

In the digit span task, participants were instructed to remember sets of digits presented in random order and recall them in ascending order. Every digit was randomly selected between 1 and 9, with the constraint that each digit could repeat across sets but not within a set. Each digit was presented for 1000 ms, with a 500 ms interval of blank screen between two successive digits. The set size started from two, and kept increasing until participants failed to correctly recall two out of three sets of a given size (n). Participants' digit span was calculated as $n-1$.

3. Results

Data of each variable was collected by following the scoring methods described in each task. For the word translation recognition task, only RTs from the critical items (i.e., 50 pairs of translation equivalents) were used for data analyses. Outliers (RTs that were three standard deviations beyond the mean RT) and RTs from incorrect trials were removed before averaging. To minimize the potential speed-accuracy trade-off effect and maintain "real" effects, the balanced integration score (BIS), which was devised to integrate speed and accuracy with equal weights (Liesefeld & Janczyk, 2019), was calculated in both E-C and C-E directions for each participant. These BISs, instead of RTs or accuracy, were used in the correlation analyses and structural equation modelling (SEM).

Data analysis was conducted in three steps. First, paired-samples t -tests were performed for all variables between pre-test (Stage 1) and post-test (Stage 2) to make a preliminary comparison. Second, Pearson correlation analyses were carried out between CI scores and other variables. Third, language skills and memory spans that were correlated with CI performance entered SEM, which was conducted to test the fitness of hypothetical models.

3.1 Descriptive statistics and paired-samples t -tests

Descriptive statistics (means and SDs) as well as the results of the paired-samples t -tests were reported in Table 1. The paired-sample t -tests showed that participants' performances in the CI tasks and in almost all the language skill tasks were significantly improved in the post-test (except for the SL summary writing in the C-E direction, and the accuracy in the word translation task which was probably due to the ceiling effect). As to the WM span tasks, a mixed pattern was revealed. A significant progress was achieved in the Chinese reading span, the English listening span and the Chinese listening span, while no significant difference was found regarding participants' performances in the English and Chinese speaking spans. What seemed weird was that participants' performance in the English reading span task was worsened at Stage 2. One possible account was that (at least some) participants, due to certain factors like fatigue, might not be as devoted to this task as they were in the pre-test, as the sentence judgement error rate in the post-test was marginally significantly higher ($t = -1.85$, $p = 0.69$). Further investigation is needed for validation.

3.2 Pearson correlation analyses

Results of Pearson correlation between E-C and C-E CI performance and other variables at Stage 1 and Stage 2 are presented in Table 2.

For E-C CI performance

As shown in the table, at Stage 1, participants' E-C CI performance correlated with 4 variables of language competence (i.e., L2 proficiency, summary writing for SL, SL listening comprehension, C-E word translation recognition BIS) and 3 WM spans (English listening span, English speaking span and Chinese speaking span). As to Stage 2, except for C-E word translation recognition BIS, all the other six variables remained to be correlated with E-C CI performance. Generally speaking, we may conclude that the variables that were correlated with E-C CI performance at Stage 1 were also correlated with the CI performance at Stage 2 during the first year of CI training.

For C-E CI performance

Only one variable (i.e., L2 proficiency) was found to be correlated with C-E CI performance at Stage 1. As to Stage 2, C-E CI performance correlated with 3 variables, i.e., summary writing for SL, SL listening comprehension and English listening span.

3.3 Structural equation modeling

To explore the self-organization of the CI competence system, we attempted to establish three hypothetical models at each stage, and examine model fitness by conducting structural equation modeling. As reviewed in "Introduction", no theoretical model of interpreting competence has specified how language competence and WM capacity may interact and then contribute together to CI performance, and only two empirical studies are relevant. Christoffels et al. (2003) examined the relationship between SI competence and two cognitive skills that are supposed to be involved in SI, i.e., WM and lexical retrieval efficiency. Results of graphical modeling indicated that both L2 reading span and L1-L2 word translation efficacy directly contributed to L2-L1 SI performance. Dong et al. (2013) tested the relationship among CI performance and potentially relevant variables with interpreting students that had received 10 months of CI training. Results of SEM revealed that interpreting students' language competence indirectly contributes to L2-L1 CI performance through the mediation of psychological competence, which included interpreting anxiety, L2 listening span and L1 speaking span. Based on such findings in the literature, we hypothesized that language competence and WM capacity were likely to have an effect on CI competence in one of the following three ways, with each model tested in Amos 24, using maximum likelihood estimation:

Hypothetical structural model A

Language competence correlates with WM capacity, and both of these two latent variables directly contribute to CI performance, as suggested by the model in Christoffels et al. (2003).

Hypothetical structural model B

WM capacity has a direct effect on CI performance, while language competence functions through the mediation of WM capacity, as suggested in Dong et al. (2013).

Table 1 Descriptive statistics (means with SD in bracket), and paired-sample t-tests for means at Stage 1 and Stage 2 in each test (N = 61).

N = 61 Task	Stage 1	Stage 2	t value	p value
<i>Consecutive interpreting tasks</i>				
E-C CI performance	61.34 (11.01)	67.21 (8.25)	-5.68	<.001
C-E CI performance	67.02 (6.01)	69.33 (5.26)	-4.21	<.001
<i>Language competence tasks</i>				
L2 proficiency	71.16 (5.40)	68.52 (5.73)		
Summary writing (E-C CI)	26.64 (16.26)	32.24 (13.59)	-2.92	.005
Summary writing (C-E CI)	63.84 (14.26)	62.84 (15.76)	.39	.700
SL listening comprehension (E-C CI)	53.48 (17.25)	63.52 (18.01)	-3.75	<.001
SL listening comprehension (C-E CI)	87.09 (13.10)	93.03 (8.70)	-3.94	<.001
C-E word translation RT	629.19 (76.07)	588.06 (71.75)	5.40	<.001
E-C word translation RT	573.40 (71.99)	549.82 (67.27)	2.88	.006
C-E word translation ACC	94.28% (3.43%)	94.30% (4.27%)	-.027	.979
E-C word translation ACC	95.03% (3.51%)	94.69% (4.27%)	.621	.537
<i>Memory span tasks</i>				
English reading span	62.43 (9.70)	59.92 (11.61)	2.23	.029
Chinese reading span	64.41 (9.17)	67.67 (6.11)	-3.38	.001
English listening span	38.82 (6.91)	40.70 (7.15)	-2.68	.010
Chinese listening span	46.62 (6.40)	50.85 (5.32)	-7.76	<.001
English speaking span	61.70 (7.56)	62.67 (8.43)	-0.98	.332
Chinese speaking span	72.28 (7.80)	72.70 (8.93)	-0.47	.642
Digit span	5.75 (.47)	5.75 (.87)	<.001	1.000

Notes: (1) Stage 1 and 2: 2nd and 10th month of interpreting training. (2) E-C: English-to-Chinese; C-E: Chinese-to-English; CI: consecutive interpreting; SL: source language. (3) Paired-sample t-test was not conducted for L2 proficiency because the data came from different tests (TEM 4 vs. TEM 8), thus were not comparable.

Table 2. Results of Pearson correlation analyses between CI performance and other variables at Stage 1 and Stage 2.

N = 61 Task	E-C CI		C-E CI	
	Stage 1	Stage 2	Stage 1	Stage 2
<i>Language competence tasks</i>				
L2 proficiency	.39**	.25*	.30*	.11
Summary writing	.34**	.36**	-.03	.29*
SL Listening comprehension	.26*	.38**	.25	.29*
C-E word translation BIS	.28*	.04	.16	.18
E-C word translation BIS	.15	-.06	.19	.08
<i>Memory span tasks</i>				
English reading span	.14	.00	-.04	-.15
Chinese reading span	.15	.20	.05	.11
English listening span	.29*	.45***	.16	.31*
Chinese listening span	.21	.12	.08	.07
English speaking span	.32*	.29*	.22	.23
Chinese speaking span	.28*	.37**	.19	.22
Digit span	.10	.15	-.08	.13

Notes: (1) Stage 1 and 2: 2nd and 10th month of interpreting training. (2) E-C: English-to-Chinese; C-E: Chinese-to-English; CI: consecutive interpreting; SL: source language; BIS: balanced integration score; (3) * $p < .05$; ** $p < .01$; *** $p < .001$

Table 3. Fit indices for the three hypothetical English-to-Chinese consecutive interpreting competence models at Stage 1 and Stage 2.

	χ^2	df	p	χ^2/df	CFI	RMSEA [90% CI]	SRMR	AIC
Cutoff Criteria			>.05	<2	>.90	<.08	<.08	The smaller, the better
Stage 1 (at the 2nd month of interpreting training)								
Model A	28.911	18	.049	1.606	.883	.101 [.005, .166]	.081	64.911
Model B	38.908	19	.005	2.048	.786	.132 [.071, .191]	.108	72.908
Model C	29.058	19	.065	1.529	.892	.094 [.000, .159]	.081	63.058
Stage 2 (at the 10th month of interpreting training)								
Model A	17.388	12	.136	1.449	.928	.087 [.000, .169]	.074	49.388
Model B	24.760	13	.025	1.905	.843	.123 [.043, .196]	.090	54.760
Model C	17.425	13	.181	1.340	.941	.075 [.000, .158]	.074	47.425

CFI: Bentler's Comparative Fit Index; RMSEA: root mean squared error of approximation; CI: confidence interval; SRMR: standardized root mean square residual; AIC: Akaike Information Criterion.

Hypothetical structural model C

Language competence directly contributes to CI performance, while WM capacity functions through the mediation of language competence.

In line with the commonly adopted guidelines for model testing (e.g., Hooper, Coughlan & Mullen, 2008; Hu & Bentler, 1999; Kline, 2016), we evaluated the fitness of each model by examining the following fit indices: chi-square (χ^2) with its degrees of freedom and p value, standardized root mean square residual (SRMR), root mean square error of approximation (RMSEA) and its 90% confidence interval (CI), Bentler Comparative Fit Index (CFI), and Akaike Information Criterion (AIC). Among these indices, the chi-square statistics were considered essential and should be reported at all times (Kline, 2016). SRMR, RMSEA and CFI were chosen over other indices, as they are more sensitive to models with misspecified factor covariances, latent structures or factor loadings (Hu & Bentler, 1999), and less affected by sample size (e.g., Hooper et al., 2008). As for AIC, it is one of the most widely accepted indices for model comparison, recommended to be used in tandem with other goodness-of-fit measures, with smaller value indicating better model parsimony. (e.g., Hooper et al., 2008).

It is generally agreed that a good-fitting model needs to meet all the following criteria: (1) a non-significant chi-square, and a ratio of chi-square to df (χ^2/df) smaller than 2 (Tabachnick & Fidell, 2007); (2) the values of the examined fit indices pass certain threshold levels. However, the cutoff values of these fit indices are not always consistent in the literature. Some researchers recommended 0.90 as the cutoff value for CFI, 0.08 for RMSEA and 0.08 for SRMR (e.g., Bentler & Bonett, 1980; Wu, 2010), while others suggested more stringent criteria (0.95 for CFI and 0.06 for RMSEA) (e.g., Hu & Bentler, 1999). Since lower cutoff values for data sets with small sample size are recommended (Browne & Cudeck, 1992; Sharma, Mukherjee, Kumar & Dillon, 2005), we decided to adopt 0.90 for CFI and 0.08 for RMSEA as the cutoff criteria, given the number of participants (N = 61) in the current study.

Results of SEM in the E-C direction at Stage 1 and Stage 2

Results of model evaluation were summarized in Table 3, with the three estimated models presented in Figure 1. A rectangle in a

model represents an observed variable measured by a specific task, while a large ellipse represents a latent variable which the observed variables have tapped. Path coefficients next to the single-headed arrows are standardized factor loadings which are equivalent to standardized regression coefficients (beta weights) estimated with maximum likelihood estimation. Squaring these loadings (i.e., number above each rectangle or ellipse) gives an estimate of the variance for each observed variable that is accounted for by the latent construct. A curved, double-headed arrow indicates a correlation between two latent variables, with its correlation coefficient placed next to the arrow.

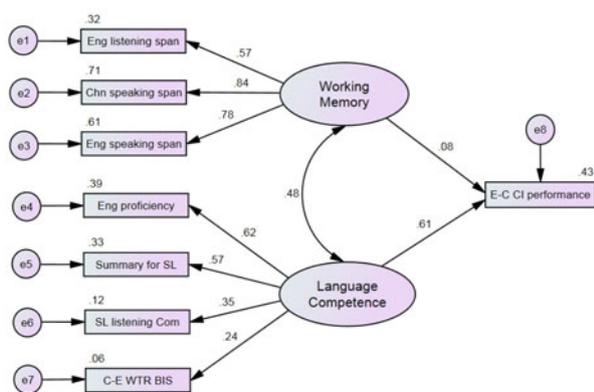
For Stage 1 (2nd month of interpreting training)

Judged by the fit indices reported in the upper part of Table 3, none of the three hypothetical models could fit the sample data, as none of their CFI, RMSEA and SRMR values passed the cutoff criteria, and (marginally) significant chi-squares were reported in all three models. Such poor fitness was consistent with the significance tests of path coefficients, the results of which indicated that each model contained at least one invalid path. [Model A: Working Memory→E-C CI performance, $\beta = .08$, $z = .452$, $p = .652$; Language Competence→C-E word translation recognition BIS, $\beta = .24$, $z = 1.456$, $p = .145$; Model B: Language Competence→SL listening comprehension, $\beta = .28$, $z = 1.584$, $p = .113$; Language Competence→C-E word translation recognition BIS, $\beta = .17$, $z = 1.007$, $p = .314$; Model C: Language Competence→C-E word translation recognition BIS, $\beta = .25$, $z = 1.558$, $p = .119$]. Therefore, no fit model of E-C CI Competence could be established at Stage 1.

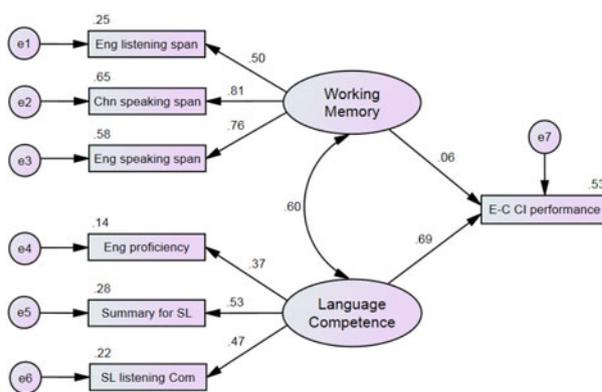
For Stage 2 (10th month of interpreting training)

The three hypothetical models constructed at Stage 2 were almost the same as those built up at Stage 1, except that the C-E word translation recognition BIS were not incorporated (because of its insignificant correlation with CI performance). Fit indices indicated that Model A and Model B could not fit the sample data, as the value of RMSEA in Model A and the values of almost all fit indices in Model B failed to pass the threshold level. Furthermore, the significance tests revealed two invalid paths in Model A [Working Memory→E-C CI performance, $\beta = .06$, $z = .206$, $p = .837$; Language Competence→E-C CI performance, $\beta = .69$, $z = 1.554$, $p = .120$], and the variance of the latent variable "language competence" in both Model A and Model B also

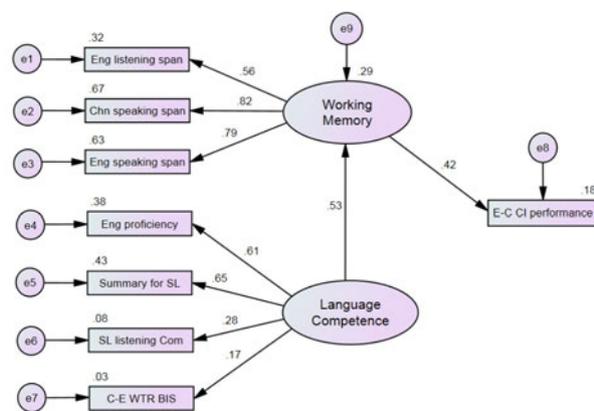
Model A at Stage 1



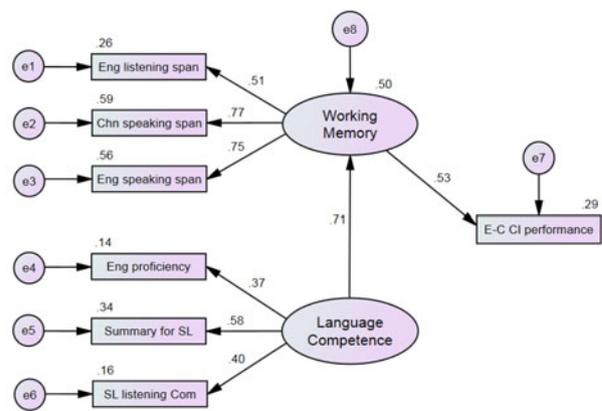
Model A at Stage 2



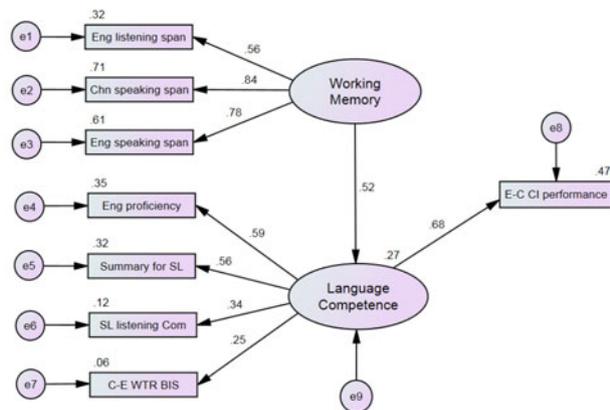
Model B at Stage 1



Model B at Stage 2



Model C at Stage 1



Model C at Stage 2

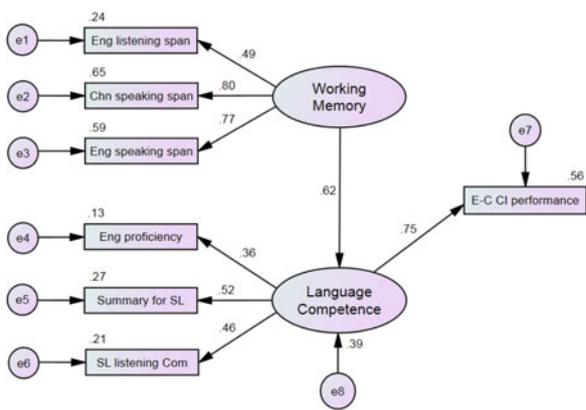


Fig. 1. The three hypothetical models of English-to-Chinese consecutive interpreting (E-C CI) competence at Stage 1 (at the left) and at Stage 2 (at the right). Only Model C at Stage 2 provided a good fit for the data. (“Summary for SL”: summary writing for source language; “SL listening Com.”: source language listening comprehension; “C-E WTR BIS”: Chinese-to-English word translation recognition balanced integration score)

failed to reach significance [Model A: $z = 1.201, p = .230$; Model B: $z = 1.128, p = .259$].

Only Model C at Stage 1 could fit the data, and thus turned out to be the accepted model. In this model, language competence directly contributed to E-C CI performance, and WM

capacity functioned through the mediation of language competence. About 39% of the variance of language competence was explained by WM, and altogether, these two latent variables accounted for about 57% of the variance of E-C CI performance.

Table 4. Fit indices for the three hypothetical Chinese-to-English consecutive interpreting competence models at Stage 2.

	χ^2	df	p	χ^2/df	CFI	RMSEA [90% CI]	SRMR	AIC
Cutoff Criteria			>.05	<2	>.90	<.08	<.08	The smaller, the better
Model A	.067	1	.795	.067	1.000	.000 [.000, .219]	.008	18.067
Model B	6.925	2	.031	3.462	.596	.203 [.052, .376]	.105	22.925
Model C	.138	2	.933	.069	1.000	.000 [.000, .071]	.012	16.138

CFI: Bentler's Comparative Fit Index; RMSEA: root mean squared error of approximation; CI: confidence interval; SRMR: standardized root mean square residual; AIC: Akaike Information Criterion.

Results of SEM in the C-E direction at Stage 1 and Stage 2

For Stage 1

Since C-E CI performance was correlated with only one observed variable, no hypothetical model could be constructed at Stage 1.

For Stage 2

Results of model evaluation for Stage 2 were summarized in Table 4, with the three estimated models presented in Figure 2. According to the fit indices reported, Model B did not fit the sample data, as a significant chi-square was reported and its values of all fit indices failed to pass the cutoff criteria. In addition, the significance tests revealed two invalid paths in Model B [Language Competence→Working Memory, $\beta = .38$, $z = .982$, $p = .326$; Language Competence→Summary for SL, $\beta = .38$, $z = .981$, $p = .327$.], and the variance of the latent variable "language competence" in Model B also failed to reach significance, $z = .840$, $p = .401$. As to Model A, although its fit indices were quite good, three paths in the model turned out to be invalid [Working Memory→C-E CI performance, $\beta = .07$, $z = .304$, $p = .761$; Language Competence→C-E CI performance, $\beta = .63$, $z = 1.263$, $p = .207$; Working Memory→Language Competence, $r = .38$, $z = 1.503$, $p = .133$.], and the variance of its latent variable "language competence" also failed to reach significance, $z = 1.150$, $p = .250$.

Again, Model C at Stage 2 turned out to be the only acceptable model, as its fit indices were excellent, and no path coefficients or variances of variables in the model failed the significance tests. The model structure was identical to the structure of the fit model for the E-C direction. About 19% of the variance of language competence was explained by WM capacity, and altogether, the two latent variables accounted for about 52% of the variance of C-E CI performance.

4. Discussion

The present longitudinal study aimed to investigate the self-organization of CI competence system in interpreting trainees. To be more specific, we intended to explore how language competence and WM, as two essential cognitive abilities recruited in interpreting, got mobilized and coordinated in interpreting students' CI competence system during the process of CI training. Two steps were taken. First, we tested how variables of language skills and memory spans correlated with CI task performances at the beginning (Stage 1) and end (Stage 2) of the first academic year of interpreting training. Second, we built and tested potential structural equation models so as to find out a fit model for the data. The results are summarized in Table 5.

4.1 Mobilization of cognitive abilities in the CI competence system

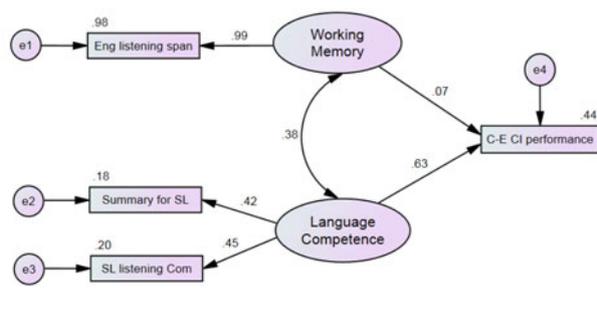
If we compare both the developmental stages and the two interpreting directions in Table 5, we may conclude that with more interpreting training, more language and memory skills become correlated with interpreting performance (at least in the first year of interpreting training as shown in the present study), suggesting the mobilization of cognitive abilities for the complex language skill of interpreting. In other words, if we mark on the time scale the less trained C-E direction before the more trained E-C direction, the developmental pattern revealed is rather neat in general: the language and memory skills that were correlated with interpreting performance at a previous stage were also correlated with interpreting performance at a later stage. There were only two exceptions. The first concerns the finding that general English proficiency was not significantly correlated with CI scores at Stage 2 in the C-E direction (see Table 5). This is probably because the students' lack of training in the C-E direction resulted in their instability in the tests, but we certainly do need more empirical studies to further verify this explanation. The second one concerns lexical retrieval efficiency (indexed by C-E word translation recognition BIS) that dropped at Stage 2 from the list of correlated factors verified at Stage 1. This is most probably due to the fact that the translation equivalent recognition task was too easy for the participants, especially at Stage 2 when they had received interpreting training for about 10 months. In a word, in spite of the two exceptions, the general pattern is clear that the development of the complex skill of interpreting was accompanied by componential cognitive abilities getting mobilized.

4.2 Emergence of organization in the CI competence system

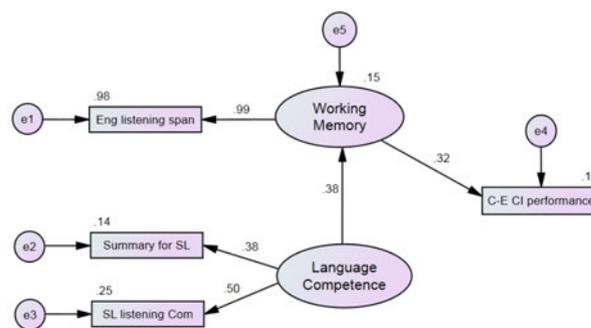
As shown in the last column in Table 5, structural equation modelling in both interpreting directions failed to yield a fit model of CI competence in the pre-test, while a fit model was consistently reported in the post-test, suggesting that the system became structured and organized after CI training. We have thus found empirical evidence showing how students' interpreting competence system evolved from an initially disorganized state to a relatively organized state.

The fit models reported in the present study give us an idea of how language competence and WM may interact within the CI competence system (at least during a certain phase). As Figure 1 and 2 show, the fit models in both E-C and C-E directions shared the same structure, indicating a relatively stable interactive pattern among these two components through which they contributed to the CI performance. Specifically, language competence directly impacted on CI performance, suggesting that better

Model A at Stage 2



Model B at Stage 2



Model C at Stage 2

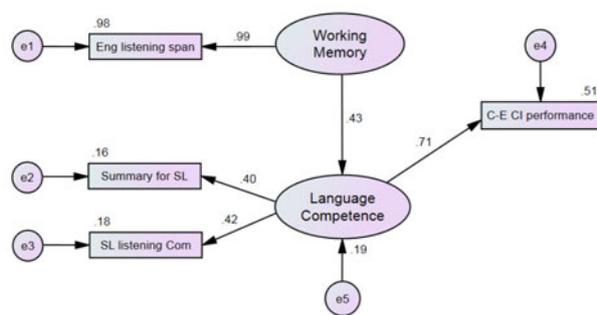


Fig. 2. The three hypothetical models of Chinese-to-English consecutive interpreting (C-E CI) competence at Stage 2, with only Model C providing a good fit for the data. (“Summary for SL”: summary writing for source language; “SL listening Com.”: source language listening comprehension)

Table 5. Number of variables correlated with consecutive interpreting performance, and whether a fit SEM model emerged at Stage 1 and Stage 2 in the two interpreting directions.

Task	Test time	Number of correlated variables	Fit model
E-C CI	Stage 1	7 (Eng proficiency, SL summary writing, SL listening comprehension, Eng listening and speaking span, Chn speaking span, C-E word translation recognition BIS)	No
	Stage 2	6 (Eng proficiency, SL summary writing, SL listening comprehension, Eng listening and speaking span, Chn speaking span)	Yes
C-E CI	Stage 1	1 (Eng proficiency)	n/a
	Stage 2	3 (SL summary writing, SL listening comprehension, Eng listening span)	Yes

Notes: Stage 1 and 2: 2nd and 10th month of interpreting training; Eng: English; SL: source language; Chn: Chinese; C-E: Chinese-to-English; BIS: balanced integration score.

language competence brought about better CI performance, while WM influenced CI performance via language competence, suggesting that people with better language competence may be able to make more effective use of their WM resources during interpreting.

The change from a disorganized state at Stage 1 to a better organized state at Stage 2 for both interpreting directions is

consistent with the process of self-organization as introduced in “Introduction”. Briefly speaking, it is the spontaneous occurrence of structured patterns, a process in which a system becomes more organized and coordinated through interactions of its components. In the specific case of interpreting, when interpreting trainees are required to complete interpreting tasks, they have to mobilize whatever abilities that are needed, and after exercising this process repeatedly, abilities that are most relevant to the successful completion of interpreting, such as key information selection (indexed by summary writing), L2 listening span, and L1 speaking span, get coordinated, giving rise to more efficient functioning of CI competence system.

5. Implications and future research

The above two major findings in the present study, when taken together, illustrate a DST perspective for the development of CI competence in interpreting students. Previous research on this issue generally viewed the development of interpreting competence as a process of transforming interpreting students’ explicit, declarative knowledge into implicit, procedural competence, which is mainly achieved in years of practice and professional experiences (Albl-Mikasa, 2013; Riccardi, 2005). Although these studies have pointed out the importance of certain influential factors, such as language competence (e.g., Cai et al., 2015), working memory (see Mellinger & Hanson, 2019 for a review), interest (e.g., Albl-Mikasa, 2013) and interpreting strategies (e.g., Dong & Li, 2020; Riccardi, 2005), some critical questions remain unclear. For instance, how do those relevant cognitive

components interact and contribute to interpreting performance? And how do their interactions change at different developmental stages? The current study provides a preliminary answer for these questions. For interpreting students, their interpreting competence develops as a result of self-organization, a process in which at least two important changes take place. First, the cognitive abilities particularly drawn upon to fulfill the specific demands of interpreting (e.g., L2 WM spans) are mobilized in training. Second, the cognitive abilities involved in interpreting tasks become better coordinated and more efficiently utilized, resulting in a phase transition from an initially disordered state to a relatively organized state. These two essential sub-processes unfold simultaneously, giving rise to the self-organization of trainees' interpreting competence system.

The self-organizing process reported in the present study helps reveal how L2 learners develop complex language skills. A complex language skill is difficult to master, as it requires efficient cooperation between different cognitive abilities. However, as the present study indicates, for a system at its initial state, it is likely that it is not well-organized and relevant cognitive abilities may fail to function in a coordinated manner. As a consequence, learners may frequently encounter difficulties, even though they may have been taught the knowledge of the skills at the very beginning. However, since language skills are open and dynamic systems, they can change continuously to adapt to new requirements. Therefore, as training goes along, the skill systems get self-organized: more cognitive abilities may be mobilized, and better coordination may be achieved among the involved cognitive abilities, leading to better development.

The present study also contributes to research on self-organization per se. Self-organization, in the framework of the DST, is particularly helpful in depicting and explaining the evolving trajectory of complex systems, which is probably why it has been studied in many fields, such as biology, physics, neuroscience, economics and cognition. Some of the research focused on the autonomy of reorganization (e.g., Shahbazi et al., 2016), while others targeted phase transition from a disorganized to a well-organized state (e.g., Kozma & Freeman, 2017), or highlighted a synergetic effect in the process of establishing order within systems (e.g., Lienen, 2014; Stadler & Kruse, 1990). As for research adopting this approach to study language processing and language development, some probed into the few essential attributes of complex dynamic systems (e.g., intra-individual variability, non-linearity and phase transition) by analyzing the characteristics of linguistic output at different stages (e.g., Baba & Nitta, 2014; de Bot et al., 2007; van Geert, 2008; Verspoor et al., 2008), while others simulated the learning process of neural networks by computational modeling (e.g., Li et al., 2007). All these studies contribute to our understanding of self-organization by inferring potential mechanisms from the input and output, but shed little light on how a cognitive system (e.g., interpreting competence) gets self-organized based on real data from components within the system. The present study directly illustrates how componential abilities get organized by using the statistic method of structural equation modelling, which has two implications for future research on self-organization. First, it is possible to directly study the self-organization of cognitive development (including language development) by exploring the relationship between its componential abilities. Second, the statistical method of structural equation modelling, a way to directly display interactions between multiple factors, may illustrate how a complex dynamic system gets organized by changes in the interactions among its componential abilities.

The present study leaves two issues for future research. First of all, replication studies are welcome to further verify the structural model yielded by SEM. The fit models of CI competence reported in the present study indicate that for interpreting trainees, their WM mainly contributes to CI performance through the mediation of language competence. This pattern was rather consistent in both E-C and C-E directions, but different from the graphic model yielded in Christoffels et al. (2003), and the structural model reported in Dong et al. (2013). In Christoffels et al. (2003), L2-L1 SI performance was directly linked to L2 reading span and L1-L2 word translation (production) efficacy, suggesting direct contribution from both language subskills and L2 WM. In Dong et al. (2013), language competence functioned through the mediation of psychological competence, which included two WM components (English listening span and Chinese speaking span) and interpreting anxiety. The distinctions in model structure might be caused by differences in the specific tasks adopted (e.g., CI vs. SI; word translation recognition vs. word translation production), in participants' background of interpreting learning history (e.g., 10 months of CI training vs. untrained), or in the categorization of individual tasks to certain latent variables (e.g., whether WM spans and interpreting anxiety are assigned to a single construct). Besides, the data sample size in the present study has just reached the bottom line for conducting structural equation modeling (5 or 10 observations per estimated parameter, Bentler & Chou, 1987), which may undermine the reliability of the SEM results to some extent. Future research with larger sample size would help further validate the interrelationship among CI competence and its cognitive components, as well as its change over time.

Second, the change of relationship between interpreting competence and other potentially related cognitive factors, such as interpreting anxiety, motivation and executive functions, awaits to be investigated. Recent studies showed that interpreting experiences may produce cognitive advantages in executive functions such as monitoring, switching and updating (see Dong & Zhong, 2019; García, Muñoz & Kogan, 2020 for reviews), suggesting that these functions may be related to interpreting tasks (see Dong & Li, 2020 for theoretical assumptions). Future longitudinal studies adopting structural equation modelling may help unveil the full picture of the relationship between interpreting competence and these cognitive factors, as well as potential changes of the relationship over time.

Acknowledgements. This work was supported by a grant to the corresponding author from the National Social Science Foundation of China [15AYY002]. We would like to thank some of the corresponding author's former graduate students who had helped in data collection, including Dr. Rendong Cai, Dr. Nan Zhao, Dr. Jiexuan Lin. No potential conflict of interest was reported by the authors.

References

- Abi-Mikasa M (2013) Developing and cultivating expert interpreter competence. *The Interpreters' Newsletter*, 17–34.
- Ashby WR (1947) Principles of the Self-Organizing Dynamic System. *The Journal of General Psychology* 37, 125–128.
- Baba K and Nitta R (2014) Phase transitions in development of writing fluency from a complex dynamic systems perspective. *Language Learning* 64, 1–35.
- Beijing Language Institute (1986) *Xiandai Hanyu Cipin Cidian [Modern Chinese Frequency Dictionary]*. Beijing: Beijing Language Institute Press.
- Bentler PM and Bonett DG (1980) Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin* 88, 588–606.

- Bentler PM and Chou C-P** (1987) Practical issues in structural modeling. *Sociological Methods & Research* **16**, 78–117.
- Browne MW and Cudeck R** (1992) Alternative ways of assessing model fit. *Sociological Methods & Research* **21**, 230–258.
- Cai R, Dong Y, Zhao N and Lin J** (2015) Factors contributing to individual differences in the development of consecutive interpreting competence for beginner student interpreters. *The Interpreter and Translator Trainer* **9**, 104–120.
- Christoffels IK and de Groot A** (2005) Simultaneous interpreting: A cognitive perspective. In Kroll JF and de Groot AMB (eds), *Handbook of bilingualism: Psycholinguistic approaches*. New York: Oxford University Press, pp. 454–479.
- Christoffels IK, de Groot AM and Kroll JF** (2006) Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language* **54**, 324–345.
- Christoffels IK, de Groot AM and Waldorp LJ** (2003) Basic skills in a complex task: A graphical model relating memory and lexical retrieval to simultaneous interpreting. *Bilingualism: Language and Cognition* **6**, 201–211.
- Daneman M and Carpenter PA** (1980) Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* **19**, 450–466.
- Daneman M and Green I** (1986) Individual differences in comprehending and producing words in context. *Journal of Memory and Language* **25**, 1–18.
- de Bot K, Lowie W and Verspoor M** (2007) A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition* **10**, 7–21.
- Dong J** (2016) A dynamic systems theory approach to development of listening strategy use and listening performance. *System* **63**, 149–165.
- Dong Y** (2018) Complex dynamic systems in students of interpreting training. *Translation and Interpreting Studies* **13**, 185–207.
- Dong Y, Cai R, Zhao N and Lin J** (2013) Xuesheng yiyuan kouyi nengli jiegou de ceshi yu fenxi [An Empirical Study on Interpreting Competence Structures in Student Interpreters]. *Waiguoyu [Journal of Foreign Languages]* **36**, 76–86.
- Dong Y and Li P** (2020) Attentional control in interpreting: A model of language control and processing control. *Bilingualism: Language and Cognition* **23**, 716–728.
- Dong Y, Li Y and Zhao N** (2019) Acquisition of interpreting strategies by student interpreters. *The Interpreter and Translator Trainer* **13**, 408–425.
- Dong Y and Zhong F** (2019) The Intense bilingual experience of interpreting and its neurocognitive consequences. In Schwieter JW (ed), *The handbook of the neurocognition of multilingualism*. US: Wiley-Blackwell, pp. 685–700.
- Elman JL, Bates EA, Johnson MH, Karmiloff-Smith A, Parisi D and Plunkett K** (1996) *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT press.
- Garcia AM, Muñoz E and Kogan B** (2020) Taxing the bilingual mind: Effects of simultaneous interpreting experience on verbal and executive mechanisms. *Bilingualism: Language and Cognition* **23**, 729–739.
- Gile D** (2009) *Basic concepts and models for interpreter and translator training (revised edition)*. Amsterdam: John Benjamins.
- Göpferich S** (2013) Translation competence: Explaining development and stagnation from a dynamic systems perspective. *Target. International Journal of Translation Studies* **25**, 61–76.
- Hooper D, Coughlan J and Mullen MR** (2008) Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods* **6**, 53–60.
- Hu L-t and Bentler PM** (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* **6**, 1–55.
- Ikeno O** (2006) L1 and L2 working memory: An investigation into the domain specificity and processing efficiency issues. *Bulletin of the Faculty of Education of the Ehime University* **53**, 113–121.
- Jin Y and Fan J** (2011) Test for English majors (TEM) in China. *Language Testing* **28**, 589–596.
- Kelso JS** (1995) *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT press.
- Kline RB** (2016) *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford publications.
- Kohonen T** (2001) *Self-organizing maps*. Berlin: Springer.
- Kozma R and Freeman WJ** (2017) Cinematic operation of the cerebral cortex interpreted via critical transitions in self-organized dynamic systems. *Frontiers in Systems Neuroscience* **11**, 1–10.
- Larsen-Freeman D** (1997) Chaos/complexity science and second language acquisition. *Applied Linguistics* **18**, 141–165.
- Larsen-Freeman D** (2006) The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics* **27**, 590–619.
- Larsen-Freeman D** (2018) Second Language Acquisition, WE, and language as a complex adaptive system (CAS). *World Englishes* **37**, 80–92.
- Lehnert G and Zimmer HD** (2008) Modality and domain specific components in auditory and visual working memory tasks. *Cognitive Processing* **9**, 53–61.
- Lewis MD** (2000) The promise of dynamic systems approaches for an integrated account of human development. *Child Development* **71**, 36–43.
- Li P, Zhao X and MacWhinney B** (2007) Dynamic self-organization and early lexical development in children. *Cognitive Science* **31**, 581–612.
- Liening A** (2014) Synergetics – Fundamental attributes of the theory of self-organization and its meaning for economics. *Modern Economy* **5**, 841–847.
- Liesefeld HR and Janczyk M** (2019) Combining speed and accuracy to control for speed-accuracy trade-offs (?) *Behavior Research Methods* **51**, 40–60.
- Liu H** (2017) *Zhiye kouyi jiaoxue yu yanjiu [Professional interpreting: Teaching, learning and researching]*. Beijing: Foreign Language Teaching and Research Press.
- Liu M, Schallert DL and Carroll PJ** (2004) Working memory and expertise in simultaneous interpreting. *Interpreting* **6**, 19–42.
- MacWhinney B and O’Grady W** (2015) *The handbook of language emergence*. US: Wiley-Blackwell.
- Mayor MJB** (2015) L2 proficiency as predictor of aptitude for interpreting: An empirical study. *Translation and Interpreting Studies* **10**, 108–132.
- Mellinger CD and Hanson TA** (2019) Meta-analyses of simultaneous interpreting and working memory. *Interpreting* **21**, 165–195.
- PACTE** (2011) Results of the validation of the PACTE translation competence model: Translation project and dynamic translation index. In O’Brien S (ed), *Cognitive Explorations of Translation*. London & New York: Bloomsbury, pp. 30–56.
- Riccardi A** (2005) On the evolution of interpreting strategies in simultaneous interpreting. *Meta: Journal des Traducteurs [Meta: Translators’ Journal]* **50**, 753–767.
- Seleskovitch D** (1999) The teaching of conference interpretation in the course of the last 50 years. *Interpreting* **4**, 55–66.
- Shahbazi MN, Jedrusik A, Vuoristo S, Recher G, Hupalowska A, Bolton V, Fogarty NM, Campbell A, Devito LG and Ilic D** (2016) Self-organization of the human embryo in the absence of maternal tissues. *Nature Cell Biology* **18**, 700–708.
- Sharma S, Mukherjee S, Kumar A and Dillon WR** (2005) A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research* **58**, 935–943.
- Stadler M and Kruse P** (1990) The self-organization perspective in cognition research: Historical remarks and new experimental approaches. In Haken H and Stadler M (eds), *Synergetics of cognition*. Springer, pp. 32–52.
- Tabachnick BG and Fidell LS** (2007) *Using multivariate statistics*. New York: Allyn and Bacon.
- Tzou Y-Z, Eslami ZR, Chen H-C and Vaid J** (2012) Effect of language proficiency and degree of formal training in simultaneous interpreting on working memory and interpreting performance: Evidence from Mandarin–English speakers. *International Journal of Bilingualism* **16**, 213–227.
- Unsworth N, Heitz RP, Schrock JC and Engle RW** (2005) An automated version of the operation span task. *Behavior Research Methods* **37**, 498–505.
- van Geert P** (2008) The dynamic systems approach in the study of L1 and L2 acquisition: An introduction. *The Modern Language Journal* **92**, 179–199.
- van Geert P** (2009) A comprehensive dynamic systems theory of language development. In De Bot K and Schrauf RW (eds),

- Language Development over the Lifespan*. New York, NY: Routledge, pp. 66–110.
- van Geert P and Verspoor M** (2015) Dynamic systems and language development. In MacWhinney B and O'Grady W (eds), *The handbook of language emergence*. US: Wiley-Blackwell, pp. 537–555.
- Verspoor M, Lowie W and van Dijk M** (2008) Variability in second language development from a dynamic systems perspective. *The Modern Language Journal* **92**, 214–231.
- Verspoor M, Schmid MS and Xu X** (2012) A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing* **21**, 239–263.
- Von Bertalanffy L** (1993) *General system theory: Foundations, development, applications*. New York: Georges Braziller, Inc.
- Wang M** (2011) Jiyu dongtai xitong lilun de wangluo yingyu yuedu moshi [A web-based English reading model-From the perspective of Dynamic Systems Theory]. *Foreign Language World*, 43–64.
- Wu M** (2010) *Jiegou fangchengmoxing – AMOS de caozuo yu yingyong* [Structural Equation modeling: Operation in AMOS and Its Application] (2 ed.). Chongqing: Chongqing University Press.
- Yu H and Lowie W** (2019) Dynamic Paths of Complexity and Accuracy in Second Language Speech: A Longitudinal Case Study of Chinese Learners. *Applied Linguistics* **41**, 855–877.