The Interpreter and Translator Trainer
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ritt20

Factors contributing to individual differences in the development of consecutive interpreting competence for beginner student interpreters

Rendong Cai[a], Yanping Dong[b], Nan Zhao[b] & Jiexuan Lin[c]
[a] School of English and Education, Guangdong University of Foreign Studies, Guangzhou, China
[b] National Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, China
[c] Faculty of English Language and Culture, Guangdong University of Foreign Studies, Guangzhou, China
Published online: 03 Mar 2015.

To cite this article: Rendong Cai, Yanping Dong, Nan Zhao & Jiexuan Lin (2015) Factors contributing to individual differences in the development of consecutive interpreting competence for beginner student interpreters, The Interpreter and Translator Trainer, 9:1, 104-120, DOI: 10.1080/1750399X.2015.1016279

To link to this article: http://dx.doi.org/10.1080/1750399X.2015.1016279

The Interpreter and Translator Trainer, 2015
Vol. 9, No. 1, 104–120, http://dx.doi.org/10.1080/1750399X.2015.1016279

Routledge
Taylor & Francis Group

Factors contributing to individual differences in the development of consecutive interpreting competence for beginner student interpreters

Rendong Cai[a], Yanping Dong[b]*, Nan Zhao[b] and Jiexuan Lin[c]

[a]*School of English and Education, Guangdong University of Foreign Studies, Guangzhou, China;* [b]*National Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, China;* [c]*Faculty of English Language and Culture, Guangdong University of Foreign Studies, Guangzhou, China*

(*Received 23 August 2013; accepted 18 July 2014*)

The present study investigates the role of memory skills (working memory and short-term memory), second language (L2) proficiency, and lexical retrieval efficiency in the development of consecutive interpreting (CI) competence. Data from 61 beginner student interpreters (unbalanced Chinese-English bilinguals) indicate that only working memory and L2 proficiency measured at the beginning of interpreting training significantly correlate with CI performances measured at both the beginning and end of the training period. As for the development of CI competence, only L2 proficiency makes a significant contribution to accounting for the variance in CI performance after removing effects of prior CI skills. The data suggest that L2 proficiency is probably the most important predictor of the development of CI competence in unbalanced beginner student interpreters and that short-term memory and working memory may play different roles in CI performance. Implications for practice in interpreting training are briefly discussed.

Keywords: consecutive interpreting development; individual differences; language proficiency; working memory; lexical retrieval

1. Introduction

Conference interpreting as extreme language use is probably one of the most complex language processing tasks possible (Obler 2012; Frauenfelder and Schriefer 1997). A natural question for those both inside and outside the field is whether interpreters possess some special abilities that allow them to interpret successfully (Russo 2011; Mackintosh 1999; Christoffels, de Groot, and Kroll 2006). This question has led to empirical efforts to identify the components of expertise in interpreting (Liu, Schallert, and Carroll 2004; Christoffels, de Groot, and Kroll 2006; Moser-Mercer et al. 2000), often defined in terms of the qualities or aptitudes that set expert interpreters apart from novice or non-interpreters (Moser-Mercer et al. 2000; Ericsson 2000). Work comparing professional interpreters to novice interpreters and untrained bilingual control groups suggests that one possible candidate for the core components of expertise is working memory (WM), as a majority of these studies reveal an interpreter advantage in this dimension (Dong and Cai, forthcoming; Köpke and Signorelli 2012).

*Corresponding author. Email: ypdong@gdufs.edu.cn

An issue related to expertise in interpreting is how to select candidates and train them to become successful interpreters. This issue has stimulated research on aptitude and aptitude testing that focuses on discovering reliable predictors of successful interpreter training (Russo 2011; Moser-Mercer 1985; Gerver et al. 1989, 1984). Research in this line usually correlates interpreting students' initial performance on tasks such as synonym generation, text recall and memory exercises, with these students' final attainment, indexed by pass rates or average scores on final examinations in interpreting. Their results suggest that memory skills may serve as a selecting criterion for interpreting candidates (Gerver et al. 1989, 1984; Moser-Mercer 1985).

Both expertise and aptitude research have enriched our understanding of interpreting competence. However, these lines of research seem to pay more attention to the final outcome than to the development of interpreting competence. One important aspect of the development of interpreting competence concerns individual differences: why do some learners learn more than others despite receiving the same amount and type of training? In other words, which factors underlie variation in the development of interpreting competence? Information about this issue has important implications, not only for theoretical models of the process and development of interpreting, but also for curriculum design for interpreter training (objectives, methods and contents, assessment procedure). This study is an attempt to work in this direction. We hope to provide initial information about the development of interpreting competence by examining factors that may affect it in a group of interpreting students in their first year of training.

2. The present study

2.1. Research question and research design

The participants in the present study were undergraduate English majors at a key foreign studies university in China. In accordance with the National College English Teaching Syllabus for English Majors (National Advisory Commission on Foreign Language Teaching in Higher Education 2000), the four-year undergraduate English major in China is divided into two stages: the foundation stage (the first and second year) and the advanced stage (the third and fourth year). At the foundation stage, all English majors receive various courses aimed at improving their basic English skills in aspects of reading, listening, speaking and writing. At the advanced stage, students begin to receive training in a selected specialisation such as translation and interpreting, linguistics and applied linguistics, or English literature. Our participants had chosen to specialise in translation and interpreting, some of them with a view to joining the profession. We collected longitudinal data pertaining to the development of their interpreting competence in the first year of training, which included four courses in interpreting and four courses in translation, together with other courses such as literature. Each course is typically 80 minutes' weekly class time for 18 weeks.

Most of the interpreting training this participant group received in the first year consisted of consecutive interpreting (CI) from L2 (the second language, English) to L1 (the first language, Chinese), as recommended by experts in conference interpretation training (Seleskovitch 1999). We focus on three possible factors underlying the development of CI competence – L2 proficiency, lexical retrieval efficiency and memory capacity – in an attempt to determine whether each of them contributes to individual differences in the development of CI competence in beginner interpreting students.

We took a longitudinal approach to addressing this question. Data were collected from the same group of participants both at the beginning (Point 1) and end of the academic year (Point 2), between which times (approx. eight months) participants received various types of formal training designed to enhance their CI competence. By statistically controlling the participant's initial CI competence,[1] we can investigate the role of different predicting factors in the improvement of their interpreting competence.

At Point 1 we measured L2 proficiency, short-term memory (STM) and WM spans, lexical retrieval speed and accuracy, and interpreting skills (based on a CI task). Then, at Point 2,[2] we once again assessed performance on the same CI task. This design allowed us to determine whether each of the three factors measured at Point 1 contributed to CI performance at Point 2 after statistically removing the effects of prior performance on this task (at Point 1).

2.2. Independent variables in the present study

The present study includes three independent variables: L2 proficiency, memory capacity and lexical retrieval efficiency. Our focus on these variables is based on both theoretical considerations and empirical evidence.

L2 proficiency

Since interpreting involves two languages, it follows that the general proficiency of both languages is important for successful interpreting. It is assumed that all of the undergraduate students in interpreting and translation have mastered their L1, their mother tongue, because in order to be accepted by a key university they had to perform well in the college entrance examination, a core subject of which is the Chinese language. However, there is likely to be large variation in these participants' L2 proficiency in the light of the results of two national English proficiency tests for English majors (Jin and Fan 2011; Dong and Lin 2013; Zou 2011; Zou and Chen 2010).

We took each participant's score on the Test for English Major Band Four (TEM4) as an indicator of L2 proficiency. TEM4, designed to test the general English proficiency of English majors in China at the intermediate level, is administered at the national level once every year and has reasonably high reliability and validity (for a detailed description of TEM 4, see Jin and Fan [2011]; Zou and Chen [2010]; Zou [2011]; Cheng [2008]).

Memory capacity

The importance of memory capacity is highlighted by the concept of 'saturation' proposed by Gile (2009), who argues that the demanding nature of interpreting often saturates the interpreter's cognitive resources, leading to interpreting errors or failures.

In cognitive psychology, multiple memory types are posited (Baddeley 1999). In interpreting studies, researchers are particularly interested in STM and WM (Köpke and Nespoulous 2006). Although the exact nature of STM and WM in neuropsychology and cognitive psychology more generally is debated (for an excellent review, see Aben, Stapert, and Blokland [2012]), in the field of interpreting studies STM generally refers to the system that temporarily stores information and WM to the system that both temporarily stores and manipulates information. Correspondingly, simple spans (e.g. digit span, word span) that mainly tap the maintenance of information (see Section 3.2.1 for a description of the tasks) are generally used to assess STM capacity, and complex spans (e.g. reading span, operation span) that tap both maintenance and manipulation of information are used to assess WM capacity.

The existence of different types of memory demands in interpreting performance is empirically supported. Christoffels, de Groot, and Waldorp (2003) find that STM and WM are both related to interpreting performance, but in different ways. Furthermore, even WM itself has a different relationship with interpreting performance depending on the task (L1 versus L2 span). These results are consistent with the finding that different kinds of memory function differently in language processing (Chincotta and Underwood 1998; Ikeno 2006; Service et al. 2002; Köpke and Nespoulous 2006; Daneman and Green 1986).

Since memory capacity plays an important role in interpreting performance, and since empirical evidence indicates that STM and WM may function differently in language processing tasks, we hypothesised that STM and WM would play different roles in the development of CI competence. Therefore, we measured both STM and WM in different tasks and examined their contribution to the development of CI competence. Specifically, we used the digit span to tap STM capacity and the listening and speaking spans in both L1 and L2 to assess WM capacity.

Lexical retrieval efficiency

Lexical retrieval efficiency refers to speed and accuracy in accessing words. Christoffels, de Groot, and Waldorp (2003) argue that efficiency in lexical retrieval, particularly the retrieval of translation equivalents, may be crucial to successful interpreting performance since the time spent finding the appropriate translation for an input word should be as short as possible given the time pressure of the interpreting task. They find that lexical retrieval efficiency indexed by word translation latency directly affects interpreting performance. Since lexical retrieval efficiency affects interpreting performance, we hypothesised that it should play a role in the development of interpreting competence. In the present study, we used a translation recognition task (de Groot and Comijs 1995; de Groot, Delmaar, and Lupker 2000) to explore the role of lexical retrieval efficiency in the development of CI competence.

3. Method

3.1. Participants

Sixty-two third-year undergraduate English majors from a key foreign studies university in China participated in the present study for a small compensation in each test they took. All participants had chosen to specialise in translation and interpreting. Due to a technical problem, data from one participant could not be recovered. Therefore, statistical analyses were based on the data from the remaining 61 participants. Participants had learned English as a foreign language at school for about 10 years and so were generally considered unbalanced bilinguals.

3.2. Procedure

All participants completed the memory, lexical retrieval and CI tasks at Point 1 and were administered the CI task again at Point 2.

3.2.1. Memory tasks

Digit span as STM. In this task, participants were required to remember combinations of digits and rearrange them in ascending order in recall. The stimuli for the digit span task were sequences of digits ranging from 1 to 9. Each digit within a given digit string was randomly selected between 1 and 9, with the constraint that each digit could repeat across sets but not within sets. The set size was presented in ascending order (i.e. from the smallest to the largest), with three trials of each set size, starting with Set Size Two (a sequence of two digits).

Procedure: Each trial started with a fixation ('+') on the centre of a computer monitor followed by the presentation of a number of digits, one at a time. The duration of the fixation and of each digit was 1000 milliseconds, with a 500 milliseconds interval between two successive digits. 500 milliseconds after the presentation of the last digit in each set, participants were instructed to recall the digits presented in ascending order by pressing the keys on the numeric keypad. Starting with the set of two digits, increasingly larger sets were presented until participants failed to recall two out of three trials of a given size. At this point, the program was terminated. The participant's digit span was the level at which s/he could correctly recall two out of three trials. To make sure that participants understood the procedure, the task started with a practice session of three trials at Set Size 2. Feedback was provided in the practice session, but not in the formal session.

English listening span as WM. The English listening span task was a listening variation of the task first developed by Daneman and Carpenter (1980, 1983).

Materials: The materials for this task comprised 60 English sentences, each of which contained 8–12 words. Of these 60 sentences, 30 were either syntactically or semantically incorrect (e.g. 'Being an environmentalist, I like to newspaper plastic bags'), with the other 30 being grammatically and semantically acceptable (e.g. 'Kate should do well in school because she is a bright child.'). To guarantee the appropriate difficulty of these sentences for the task, we administered a norming test with a separate group of 26 participants from the same population as those in the main experiment. In the norming test, participants were required to listen to 90 sentences under normal listening conditions (i.e. a single task without additional recalling requirement) and judge whether each sentence they heard made sense. On the basis of the results of this test, we chose 60 sentences for which accuracy was above 90%.

Procedure: Participants listened through earphones to sentences presented via E-Prime software while trying to remember the last word of each sentence. For example, in the following set of three sentences, participants listened to one sentence at a time and then judged whether the sentence made sense via key press (the F or J key on the keyboard).

Being an environmentalist, I like to newspaper plastic bags.

Kate should do well in school because she is a bright child.

Campus violence has been rising in many cities in China.

After the last sentence in each set, participants received an auditory and visual cue to recall the last word of each sentence presented. They were instructed not to start recalling from the last word of the last sentence of each set. Apart from that, no other restrictions in the order of recall were imposed. The number of sentences (set size) varied between two and six, with three trials of each set size. The order of set size presentation was randomised across participants. Following the conventional practice in the field (Hannon and Daneman 2001; Daneman and Hannon 2001; Kane et al. 2004; Engle et al. 1999; Turner and Engle 1989), we did not take accuracy into consideration when calculating the span score. Thus, the listening span score was the cumulative number of words recalled from all trials (maximum score = 60).

Chinese listening span as WM. The stimuli of the Chinese listening span task consisted of 60 Chinese sentences, 30 of which are normal sentences (e.g. '在与人交谈时，要专注，积极倾听他的谈话，不时地给予适当的反应 [When talking to someone, be focused and attentive to his words, and give appropriate responses from time to time]') and the other 30 are either syntactically or semantically incorrect (e.g. '很少人知道一天之中最冷的时候，不是番茄炒蛋，而是在临近黎明 [Few people know that the coldest time of day is not tomato omelette, but close to dawn]'). Again, a norming test was conducted to guarantee the appropriateness of the sentences used in the task. The norming test involved recruiting a separate group of 26 participants to listen to 90 sentences under normal listening conditions and to make veracity judgments. We chose 60 sentences for the present task on the basis of the results of the norming test.

Procedure: Apart from the stimuli being Chinese sentences instead of English sentences, the procedure for the Chinese listening span task was identical to that for the English listening span task.

English speaking span as WM. The English speaking span task was based on the paradigm originally developed by Daneman and her colleague (Daneman 1991; Daneman and Green 1986).

Materials: The stimuli were 100 English words comparable in phonetic and orthographic length, containing two syllables and seven to eight letters. Most of the words were taken from Mizera (2006). Although Mizera's (2006) study was also about L2 WM, to ensure the appropriate difficulty of the materials for the participants of the present study we conducted a norming test with a separate group of 20 participants from the same population as those in the main experiment. The material for the norming test consisted of 130 words, the original 100 words from Mizera (2006) and 30 new words. All new words were matched with the original 100 words on orthographic and phonetic length. In the norming test, participants were asked to mark words they were familiar with on a list of 130 words. The results showed that 105 out of 130 words were familiar to all 20 participants, indicating that most of the 130 words were easy for them. On the basis of these results, we randomly chose 100 out of the 105 familiar words as the stimuli for the English speaking span task.

Procedure: First, participants read the instructions for the task and confirmed that they understood the procedure before actually beginning. During the task, participants attempted to remember the words presented on the centre of a computer screen and then to use those words to make sentences. For example, in a set of three words (e.g. 'explain, disturb, picture'), participants were shown one word at a time on the screen. After a short alert, a word appeared at the centre of the screen for 1000 milliseconds. After 500 milliseconds the next word was presented. The end of the set was signalled visually on the computer screen and by another sound. After the presentation of a complete set, the participant was asked to generate aloud sentences containing the presented words. The total number of correctly recalled words in the grammatically correct sentences was defined as the participant's English speaking span (maximum score = 100).

Chinese speaking span as WM. Apart from the stimuli being two-character Chinese words instead of English words, the design, procedure and scoring method for the Chinese speaking span task was the same as that for the English speaking span task. All stimuli for the Chinese speaking span task were high frequency words, taken from Table Four (1) in *Modern Chinese Frequency Dictionary* (Beijing Language Institute 1986).

3.2.2. *Word translation recognition task*

Materials: The stimuli were 200 Chinese and 200 English words, from which two lists of word pairs were compiled. One list was used in the translation recognition task in the direction of Chinese to English ('C-E translation'), and the other in the direction of English to Chinese ('E-C translation'). Each list consisted of 100 word pairs, 50 pairs of translation equivalents and 50 pairs of fillers that were not translations equivalent of each other. The 50 pairs of translation equivalents were critical items that were entered into statistical analyses.[3] The two lists were presented to participants in random order. In addition, 12 word pairs were compiled for practice trials.

Procedure: The computer-based word-translation recognition task was programmed with E-prime software. The task consisted of two sessions. One session was the C-E translation task, in which a Chinese word appeared prior to an English word for each word pair. The other session was the E-C translation task, in which an English word preceded a Chinese word for each word pair. The order of the two sessions was randomised across participants by the E-prime program to avoid a potential effect of the order of the translation directions. As a result, participants either completed the C-E translation session and then the E-C translation session or vice versa.

Before the session started, participants read the task instructions. The session began with a practice block of 10 items. Following this practice block was the experimental block of 100 items, preceded by 2 more practice items. The procedure for the practice block and the experimental block was the same. On each trial, a fixation stimulus ('+') first appeared at the centre of the screen for one second. On its disappearance, the first word was presented on the screen at a position slightly above the centre of the screen. Fifty milliseconds after the appearance of this first word, the other word in the pair appeared below it. At this point, participants indicate via key press whether the second word was the translation equivalent of the first word. They were instructed to press the 'F' key if the answer was 'yes' and the 'J' key if the answer was 'no'. Accuracy and response time (RT) data were both registered by the computer. Following the key press, the two words simultaneously disappeared. For practice trials, the disappearance of the words was immediately followed by presentation of accuracy and RT feedback for 500 ms; after a 500 ms interval of blank screen a new trial would start. However, for the trials in the experimental block, the words' disappearance was not followed by any feedback, but rather by an interval of 500 milliseconds and then the next trial. Feedback on average accuracy and RT was provided when all trials in the block were finished. After the first session, participants took a short break and then pressed the space bar to enter the next session.

3.2.3. *Consecutive interpreting task*

In the CI task, participants were required to orally translate the source language (English) speech into the target language (Chinese). All stimuli were presented aurally to the participants over headphones and all spoken responses were recorded on the computer.

Materials: The E-C CI task comprised an approximately 8-minute-long recording of a speech made at a promotion for laptops. The speech was given by a native English-speaking male at an average rate of 143 words per minute. We divided the speech into segments, with each segment consisting of two to three sentences. This is rather short in relation to general CI training practice in Europe which uses segments of around six minutes. Our segmentation is based on the fact that our participants were unbalanced Chinese-English bilinguals who had just started interpreting training, and longer segments would be too demanding for them.[4] To confirm that the level of difficulty was appropriate for our participants, we conducted a pilot study with a separate group of 20 participants. This group was sampled from the same population as the participants of the main study so they had a similar level of English proficiency and interpreting competence. In the pilot study, participants were required to translate the English speech into Chinese and then complete a questionnaire about the difficulty of the task. Furthermore, we asked five interpreting teachers from the same university and with years of experience in the field to evaluate the difficulty of the task. Both the participants in the pilot study and the interpreting teachers rated the materials as appropriate for the participants in the main study. Finally, participants in the main experiment, after completing the CI task, were also asked to report whether the difficulty level was appropriate, but no complaints about the difficulty level were made.

Procedure: The E-C CI task was administered in groups in an interpreting training lab at the university where the participants had been recruited. First, detailed instructions were provided on paper and administered verbally to participants. All participants confirmed that they understood the procedure of the task before they began the main part (i.e. interpreting). In the main part, participants listened to each segment of the speech one at a time; for example, 'Now I'm gonna use the rest of my time to talk about One-Laptop-Per-Child. But think of it not as a laptop project; the laptop is almost irrelevant. It's an education project.' The end of each segment was signalled by a sound. The sound also served as a cue that participants should start to interpret. The duration allowed for interpreting the segment was 1.5 times the duration of the segment itself, a constraint based on our pilot work. The computers recorded what participants uttered aloud during the interpreting period. Completion of the interpreting period was signalled by another sound. After a brief interval, a new segment was presented; for example, 'Because this was the first project that really had laptops at scale, eh, in a very remote village, in fact in a village that has no electricity, no telephone, no television … And there are four other villages involved, two of which have no road.' Participants were allowed to take notes and refer back to these notes as they completed the task.

Scoring: Two judges were asked to listen to the recordings and to rate each participant's performance. Both judges were interpreting teachers and were themselves professional interpreters with many years of experience in the field. The criteria for assessing the CI quality included information (accuracy and completeness) and target language (grammar and appropriateness). The inter-rater coefficient for the rating was 0.89.

4. Results

Means and standard deviations for all the variables of the present study are displayed in Table 1. The correlation and regression analyses, described next, are of primary interest. However, for completeness we also performed three paired-samples *t*-tests[5] to examine whether language had an effect on the performance on the WM span and word-translation tasks.

Table 1. Means and standard deviations (SD) for all variables (n = 61).

Variable	Mean	SD
English proficiency	71.16	5.4
Digit span	5.75	0.47
Chinese listening span	46.62	6.4
English listening span	38.82	6.91
Chinese speaking span	72.28	7.8
English speaking span	61.7	7.56
C-E word translation (RT)	629.19	76.07
E-C word translation (RT)	573.4	71.99
C-E word translation (accuracy)	0.94	0.03
E-C word translation (accuracy)	0.95	0.04
Consecutive interpreting (Point 1)	61.34	11.01
Consecutive interpreting (Point 2)	67.21	8.25

Note: For English proficiency, the maximum score is 100; for digit span, the reported average is the number of correctly recalled digits (maximum score = 9); for listening and speaking spans, the number of correctly recalled words (maximum score = 60 for listening span, 100 for speaking span); for word translation, the reaction time (RT) for correct response in milliseconds and the accuracy of response; for consecutive interpreting, the meaning score of two raters (maximum score = 100).

The results indicated that the differences between the two language versions of each WM span task reached significance: listening span, $t(60) = 10.959$, $p = .00$; speaking span, $t(60) = 13.093$, $p = .00$, with the average score for L1 significantly higher than that for L2. This is consistent with previous studies (Chincotta and Underwood 1998; Service et al. 2002) in which less advanced bilinguals manifested an L1 over L2 WM span advantage greater than more advanced bilinguals. Our participants' L1 dominance was confirmed by the direction effect in word translation as revealed by a paired-sampled *t*-test with the RT data from the word translation task (t = 9.527, df = 60, p = .00): Our participants responded significantly faster from L2 to L1 than from L1 to L2. This was also consistent with findings from word translation studies (Kroll and Stewart 1994; Kroll et al. 2002).

The correlations[6] between each predicting variable and CI performance at each test time are presented in Table 2. English proficiency significantly correlated with CI performance at both Point 1 and Point 2, suggesting that for beginner interpreters, L2

Table 2. Correlation between performance on all predicting variables and performance on the consecutive interpreting (CI) task at two times.

Predicting variable	CI (Point 1)	CI (Point 2)
English proficiency	.387**	.439**
Digit span	0.098	0.153
Chinese listening span	0.207	0.197
English listening span	.287*	.303*
Chinese speaking span	.275*	0.225
English speaking span	.319*	.327*
C-E word translation (RT)	−0.115	−0.184
E-C word translation (RT)	0.033	−0.099
C-E word translation (accuracy)	0.227	0.08
E-C word translation (accuracy)	0.243	0.135

Note: *$P < .05$. **$P < .01$.

proficiency plays a significant role in CI performance. In contrast, none of the correlations between measures of word translation and CI performance reached significance, which suggests that for these participants the speed and accuracy of lexical retrieval did not make a significant contribution to their CI performance.

As for the relationship between memory skills and CI performance, the pattern is more complicated. Firstly, Table 2 shows that the digit span and WM spans have different relations with CI performance. The correlations between digit span and CI performance did not reach significance. In contrast, both English listening and speaking spans significantly correlated with CI performance. This difference suggests that STM and WM function differently in the CI task. Furthermore, the pattern of correlations between WM spans and CI performance was somewhat different for WM spans measured in L1 versus L2. Listening span measured in English, but not in Chinese, significantly correlated with CI performance; speaking span measured in English significantly correlated with CI performance at both Point 1 and Point 2 whereas the task measured in Chinese correlated with Point 1 CI performance only. In addition, at Point 1, the numerical strength of the relationship between English speaking span and CI was stronger than that between Chinese speaking span and CI. The pattern of correlations between WM spans and CI performance indicates that WM capacity measured in different languages played somewhat different roles in CI.

One important observation from Table 2 is that each English predictor (English proficiency English listening span, and English speaking span) significantly predicts variance in E-C CI performance at both Point 1 and Point 2. Since these variables predict performance at both times, we might ask whether they predict variance in the improvement of CI competence.[7] To examine this issue, we need to control for initial interpreting competence and then examine the relationship between the predicting variables and performance at Point 2. To achieve this, we performed a series of stepwise regression analyses,[8] in which the scores on E-C CI at Point 1 were entered into the regression equation first to statistically control for their influence on E-C CI performance at Point 2. After that, in one analysis (the result of which is presented in Table 3, Model 1), we entered the English listening span at Point 1 as the second predictor to test whether it contributed, above and beyond that of E-C CI at Point 1, to E-C CI performance at Point 2; in the second analysis (Model 2), we instead entered English speaking span as the second predictor; and in the third analysis, general English proficiency as the second predictor (Model 3).

Table 3. Summary of hierarchical multiple regression analysis for variables predicting consecutive interpreting (CI) performance at Point 2.

Variable	R	R2	ΔR2	ΔF
	Model 1			
Point 1 CI	0.682	0.466	0.466	51.445**
English listening span	0.692	0.478	0.012	1.381
	Model 2			
Point 1 CI	0.682	0.466	0.466	51.445**
English speaking span	0.692	0.479	0.013	1.478
	Model 3			
Point 1 CI	0.682	0.466	0.466	51.445**
English proficiency	0.708	0.502	0.036	4.170*

Note: $*P < .05$. $**P < .01$.

As shown in Table 3, after partialling out the contribution of E-C CI performance at Point 1, neither of the WM spans at Point 1 made a statistically meaningful contribution to the variance E-C CI performance at Point 2. This suggests that both spans have limited value in predicting gains in competence in the E-C CI task, although the English listening span, English speaking span and Chinese speaking span respectively accounted for 8.3%, 10.2% and 7.5% of the variance in E-C CI performance at Point 1 (see Table 2), which in turn accounted for 46.6% of the variance in the same task at Point 2 (see Table 3). However, unlike WM spans, after partialling out the influence of prior E-C CI skills, English proficiency still accounted for an additional 3.6% of the variance in E-C CI performance at Point 2, which was statistically meaningful. This suggests that L2 proficiency did play a significant role in the development of E-C CI competence for our participants.

5. Discussion

The purpose of this study was to examine factors affecting the development of CI competence at a very initial stage. We focused on three factors: memory skills, lexical retrieval efficiency and L2 proficiency. The major findings are as follows: (1) English proficiency significantly correlated with CI performance at both points in time; (2) none of the measures of lexical efficiency had a statistically meaningful relationship with CI scores; (3) the digit span and WM spans had different relationships with CI performance; furthermore, WM spans measured in different languages also had a different relationship with CI performance; and (4) although English proficiency, English listening span, and speaking spans, can each predict CI performance at both Point 1 and Point 2, only English proficiency can make a unique contribution in accounting for variance in the development of CI competence. We will address each of these findings in turn.

Lexical retrieval efficiency and CI. Measures of lexical retrieval efficiency in the present study failed to make a meaningful contribution to accounting for variance in interpreting performance. This result was different from that of Christoffels, de Groot, and Waldorp (2003) who discovered that lexical retrieval efficiency is an important contributor to interpreting performance. This discrepancy may stem from both participant characteristics and the tasks used. The participants in the present study were unbalanced bilinguals whose L2 proficiency was relatively low, as suggested by the scores of the CI task and, importantly, the direction effect in word translation. Compared to our participants, the participants in Christoffels, de Groot, and Waldorp (2003) were relatively more proficient in L2, as indicated by their self-rating scores and the null direction effect in word translation in their study. We used the translation recognition task to measure lexical retrieval efficiency and Christoffels, de Groot, and Waldorp (2003) used the translation production task to tap lexical retrieval efficiency. Our task measured mainly retrieval efficiency; Christoffels et al.'s task measured not only retrieval of the translation equivalents, but also the production of translation equivalents, which bears a closer resemblance to the interpreting task.

Unlike Christoffels, de Groot, and Waldorp (2003), who find that lexical retrieval efficiency is related to simultaneous interpreting (SI) performance, we correlated lexical retrieval efficiency with CI performance. Compared with SI, CI imposes less time constraints on interpreters (Gile [1997] 2002), meaning that our participants had more time to retrieve translation equivalents relative to participants in Christoffels, de Groot, and Waldorp (2003). This in turn might contribute to the failure to observe a meaningful relationship between lexical retrieval efficiency and CI performance in the present study.

Memory skills and CI. In contrast to the non-significant correlation between the digit span (i.e. STM) and CI performance at both Point 1 and Point 2, three out of the four WM spans significantly correlated with CI performance at one or both of these time points. This pattern of results is in line with similar studies (e.g. Köpke and Nespoulous 2006; Christoffels, de Groot, and Waldorp 2003) and may be attributed to the functional difference between STM and WM (Engle et al. 1999; Aben, Stapert, and Blokland 2012). The digit span taps STM capacity whose function mainly lies in passive storage, whereas the listening and speaking spans tap not only storage, but also processing components. Language processing, especially that involved in a task as complex as interpreting, relies on both storage and processing functions of memory. This may explain why the WM spans rather than the STM span related to CI performance.

Furthermore, we observed a language effect on the relationship between WM and CI performance. Specifically, the WM spans measured in L2 had a stronger relationship with CI performance compared with the WM spans measured in L1. This result is consistent with the domain-specific view of WM. In WM studies, some researchers propose the domain-specificity of WM resources in the sense that different domains of processing require different types of cognitive resources (Shah and Miyake 1996; Daneman and Tardif 1987; Ketelsen and Welsh 2010). Therefore, it is possible that L1 and L2 processing require different pools of cognitive resources; empirical studies (Ikeno 2006; Alptekin and Erçetin 2010) report that L1 and L2 WM spans have different predictive power for L2 comprehension. In addition, researchers in WM studies (Shah and Miyake 1996; Just and Carpenter 1992) also propose that WM effects manifest themselves mainly in capacity-demanding tasks. In the context of the present study, although the task of E-C CI obviously requires both L1 and L2 cognitive resources since CI involves the processing of both languages, the processing of L2 was relatively more demanding. First, for the CI task, as pointed out by Gile ([1997]2002), the more difficult component lies in the comprehension of SL (in this case, English) relative to the production of the target language (in this case, Chinese); and this relative difficulty was boosted by the imbalance of the participants' two languages. Taken together, the domain-specificity of WM and the relatively more capacity-demanding nature of comprehension might underlie the difference in the relationship between WM in L1 or L2 with CI performance.

L2 proficiency and CI. In the present study, English proficiency significantly correlated with CI performance at both Point 1 and Point 2, suggesting that for beginner interpreters, L2 proficiency plays a significant role in CI performance. This pattern of results replicates that of a study that also examined the role of L2 proficiency in interpreting performance in Chinese-English bilinguals (Tzou et al. 2012). Tzou et al.'s participants were professional interpreters and relatively advanced student interpreters (at the graduate level) whereas our participants were beginner student interpreters. Taken together, it appears that L2 proficiency not only plays a role in interpreting performance for beginner student interpreters but also for advanced student interpreters, and professionals.

One of the main motivations for the present study was to explore factors affecting the development of CI competence. Although results from our correlational analyses suggest that both WM spans and English proficiency may account for variance in the performance of CI at both Point 1 and Point 2, results from our hierarchical multiple regression analyses show that only L2 proficiency can predict variance in the *development* of CI competence. This contrast suggests that for the development of CI competence at the beginner stage, L2 proficiency seems to play a more important role than WM or lexical retrieval efficiency. Of course, it should be noted that the results of the present study do not rule out the importance of cognitive resources or lexical retrieval efficiency in CI development. It is possible that these two factors exert their influence at a later, but not at an initial, stage. In addition, the segment of the CI task was relatively short and participants were allowed to take notes during interpreting, which might exert an influence on the results. With more balanced bilinguals and longer segmentation of the input speech, the results may have included a contribution of WM to CI development.

Another point that should be noted is the relatively small proportion of the unique contribution made by L2 proficiency to the variance of the development of CI competence. According to Table 3, in addition to those explained at Point 1 CI (46.6%), L2 proficiency can uniquely explain only 3.6% of the variance observed. However, recall from Table 2 that L2 proficiency significantly correlated with Point 1 CI and a further regression analysis showed that it was able to explain 15% of the variance in the performance on the Point 1 CI task. Therefore, in addition to a direct effect on CI performance at Point 2, L2 proficiency had an indirect effect on Point 2 CI mediated by Point 1 CI.

This finding has an implication for CI training: at the beginner stage, learners may benefit more when instructors design exercises to improve their L2 proficiency compared with exercises to improve their memory capacity or lexical retrieval efficiency (for a similar suggestion deriving from a different perspective, see Abuín González [2012]). Another implication is that the best candidates for interpreting training are those who have mastered their L2, which is easy to understand since it is in line with almost everyone's intuition.[9]

The present study is only the first step to discovering predictors of the development of interpreting competence. A more complete picture requires empirical research with interpreting trainees from different backgrounds, especially different L2 proficiency levels, and with more tasks or experimental manipulations related to interpreting trainees' expertise and aptitude.

Acknowledgements

We thank Dr Kalim Gonzales and two anonymous reviewers for their insightful comments on and helpful suggestions for previous drafts.

Funding

This research was supported by grants to the corresponding author from the National Social Science Foundation of China [grant number10BYY010]; the Chinese Ministry of Education [grant number 2009JJD740007]; and the New Century Talents Program by the Chinese Ministry of Education.

Notes

1. We used stepwise regression analysis to statistically control for the effect of the participants' initial CI competence. Stepwise regression analysis is often used in behavioural studies to investigate the relationship between a set of independent variables and a dependent variable. By varying the order of independent variables entering into the regression equation, the stepwise regression analysis can determine the unique contribution made by an independent variable in accounting for the variance in the dependent variable. See the Results section for the details of this analysis.

2. The participants performed the same CI task twice. Although there may be concern about a potential practice effect, we reduced this potential effect to the minimum in our test administration. First, the interval between the two tests was almost two semesters, which was much longer than many developmental studies in the literature (e.g. one-week interval in Quirmbach et al. [2009], six-week interval in Dalton [2011] and eight-week interval in Mackay, McCluskey, and Mayes [2010]). Second, the participants were not aware of the purpose of the study and they did not know that they would perform the same test again later. Therefore, they were not motivated to remember anything specific; this was confirmed in interviews with participants conducted after the tests.
3. The critical translation pairs were selected on the basis of their baseline response times in a lexical decision task in a preparatory study. To collect the baseline response times for these words, we recruited 23 participants from the same population with the participants of the main study to perform a lexical decision task. From this preparatory study, the response times of a set of 300 words (150 Chinese words and 150 English words) were collected. Out of these 300 words, we selected 100 words (50 Chinese words and 50 English words) whose response times were relatively closer to each other indicated by the small standard deviations relative to the means (Chinese: mean = 544.41, SD = 18.29, range = 61.29, minimum = 519.83, maximum = 581.11; English: mean = 561.69, SD = 25.64, range = 100.25, minimum = 499.28, maximum = 599.53).
4. In fact, the interpreting task in the present study was already challenging enough. Although the participants were allowed to take notes during the task, and the difficulty level of the speech was reported as appropriate by both the participants and their teachers, the mean score for the interpreting product was about 61.34 out 100 at Point 1 and 67.21 out 100 at Point 2.
5. A *t*-test is used to determine whether the difference between two sets of data is significant. If the result of *t*-test reaches the significant level ($p < .05$), it is said that the two sets of data are significantly different from each other.
6. The correlation coefficient is a quantitative measure of the extent to which two variables are related and the direction of that relationship. Two variables are positively correlated if a high score on one variable is related to a high score on the other, and two variables are negatively correlated if a high score on one variable is related to a low score in the other.
7. Statistic analysis indicated that our participants improved their CI competence during the interval between Point 1 and Point 2: a paired-samples *t*-test was performed on the data of the CI task at both points in time and the result was that the mean score at Point 2 was significantly higher than that at Point 1 ($t = 5.683$, $df = 60$, $p = .00$), suggesting that our participants improved their CI skills after one year's training.
8. See note 1 for the statistical meaning of the stepwise regression analysis.
9. As pointed out by Dr Kalim Gonzales (via email), it is also possible that the predictive power of L2 proficiency on CI performance at Point 2 could simply reflect the fact that students who are more motivated to become proficient in English are likely to be the same students more motivated to succeed at CI training (and thus to study harder in the eight-month academic year between P1 and P2). This possibility, however, does rule out the implication that the best candidates for interpreting training are those who have mastered their L2.

References

Aben, B., S. Stapert, and A. Blokland. 2012. "About the Distinction between Working Memory and Short-Term Memory." *Frontiers in Psychology* 3. doi:10.3389/fpsyg.2012.00301.

Abuín González, M. 2012. "The Language of Consecutive Interpreters' Notes: Differences Across Levels of Expertise." *Interpreting* 14 (1): 55–72. doi:10.1075/intp.14.1.03abu.

Alptekin, C., and G. Erçetin. 2010. "The Role of L1 and L2 Working Memory in Literal and Inferential Comprehension in L2 Reading." *Journal of Research in Reading* 33 (2): 206–219. doi:10.1111/j.1467-9817.2009.01412.x.

Baddeley, A. D. 1999. *Essentials of Human Memory.* East Sussex: Psychology Press.

Beijing Language Institute. 1986. *Xiandai Hanyu Cipin Cidian* [Modern Chinese Frequency Dictionary]. Beijing: Beijing Language Institute Press.

Cheng, L. 2008. "The Key to Success: English Language Testing in China." *Language Testing* 25 (1): 15–37. doi:10.1177/0265532207083743.

Chincotta, D., and G. Underwood. 1998. "Simultaneous Interpreters and the Effect of Concurrent Articulation on Immediate Memory: A Bilingual Digit Span Study." *Interpreting* 3 (1): 1–20. doi:10.1075/intp.3.1.01chi.

Christoffels, I. K., A. M. B. de Groot, and J. F. Kroll. 2006. "Memory and Language Skills in Simultaneous Interpreters: The Role of Expertise and Language Proficiency." *Journal of Memory and Language* 54 (3): 324–345. doi:10.1016/j.jml.2005.12.004.

Christoffels, I. K., A. M. B. de Groot, and L. J. Waldorp. 2003. "Basic Skills in A Complex Task: A Graphical Model Relating Memory and Lexical Retrieval to Simultaneous Interpreting." *Bilingualism: Language and Cognition* 6 (3): 201–211. doi:10.1017/S1366728903001135.

Dalton, T. A. 2011. "Comparison of Two Approaches to Improving Cognitive Academic Language Proficiency for School-Aged, English Language Learners: Two-Group, Pretest/Posttest Design." *All Graduate Reports and Creative Projects*. *Paper 41* http://digitalcommons.usu.edu/gradreports/41.

Daneman, M. 1991. "Working Memory as a Predictor of Verbal Fluency." *Journal of Psycholinguistic Research* 20 (6): 445–464. doi:10.1007/bf01067637.

Daneman, M., and P. A. Carpenter. 1980. "Individual Differences in Working Memory and Reading." *Journal of Verbal Learning and Verbal Behavior* 19: 450–466. doi:10.1016/S0022-5371(80)90312-6.

Daneman, M., and P. A. Carpenter. 1983. "Individual Differences in Integrating Information between and within Sentences." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9 (4): 561–584. doi:10.1037/0278-7393.9.4.561.

Daneman, M., and I. Green. 1986. "Individual Differences in Comprehending and Producing Words in Context." *Journal of Memory and Language* 25 (1): 1–18. doi:10.1016/0749-596x(86)90018-5.

Daneman, M., and B. Hannon. 2001. "Using Working Memory Theory to Investigate the Construct Validity of Multiple-Choice Reading Comprehension Tests Such as the SAT." *Journal of Experimental Psychology: General* 130 (2): 208–223. doi:10.1037/0096-3445.130.2.208.

Daneman, M., and T. Tardif. 1987. "Working Memory and Reading Skill Re-Examined." In *Attention and Performance XII*, edited by M. Coltheart, 491–508. London: Erlbaum.

De Groot, A. M. B., and H. Comijs. 1995. "Translation Recognition and Translation Production: Comparing a New and an Old Tool in the Study of Bilingualism." *Language Learning* 45 (3): 467–509. doi:10.1111/j.1467-1770.1995.tb00449.x.

De Groot, A. M. B., P. Delmaar, and S. J. Lupker. 2000. "The Processing of Interlexical Homographs in Translation Recognition and Lexical Decision: Support for Non-Selective Access to Bilingual Memory." *Quarterly Journal of Experimental Psychology* 53A (2): 397–428. doi:10.1080/713755891.

Dong, Y., and R. Cai. 2015. "Working Memory in Interpreting." In *Working Memory in Second Language Acquisition and Processing: Theory, Research and Commentary*, edited by Z. Wen, M. Mota, and A. McNeill, 64–82. Bristol: Multilingual Matters.

Dong, Y., and J. Lin. 2013. "Parallel Processing of the Target Language during Source Language Comprehension in Interpreting." *Bilingualism: Language and Cognition* 16 (03): 682–692. doi: doi:10.1017/S1366728913000102.

Engle, R. W., S. W. Tuholski, J. E. Laughlin, and A. R. A. Conway. 1999. "Working Memory, Short-Term Memory, and General Fluid Intelligence: A Latent-Variable Approach." *Journal of Experimental Psychology: General* 128 (3): 309–331. doi:10.1037/0096-3445.128.3.309.

Ericsson, K. A. 2000. "Expertise in Interpreting: An Expert-Performance Perspective." *Interpreting* 5 (2): 187–220. doi:10.1075/intp.5.2.08eri.

Frauenfelder, U. H., and H. Schriefer. 1997. "A Psycholinguistic Perspective on Simultaneous Interpretation." *Interpreting* 2: 55–89. doi:10.1075/intp.2.1-2.03fra.

Gerver, D., P. Longley, E, J. Long, and S. Lambert. 1984. "Selecting Trainee Conference Interpreters: A Preliminary Study." *Journal of Occupational Psychology* 57 (1): 17–31. doi:10.1111/j.2044-8325.1984.tb00144.x.

Gerver, D., P. Longley, E, J. Long, and S. Lambert. 1989. "Selection Tests for Trainee Conference Interpreters." *Méta* 34 (4): 724–735.

Gile, D. 2009. *Basic Concepts and Models for Interpreter and Translator Training (Revised Version)*. Amsterdam/Philadelphia: John Benjamins.

Gile, D. [1997]2002. "Conference Interpreting as a Cognitive Management Problem." In *The Interpreting Studies Reader*, edited by F. Pöchhacker and M. Shlesinger, 162–176. London: Routledge.
Hannon, B., and M. Daneman. 2001. "A New Tool for Measuring and Understanding Individual Differences in the Component Processes of Reading Comprehension." *Journal of Educational Psychology* 93 (1): 103–128. doi:10.1037/0022-0663.93.1.103.
Ikeno, O. 2006. "L1 and L2 Working Memory: An Investigation into the Domain Specificity and Processing Efficiency Issues." *Bulletin of Ehime Universities* 53 (1): 113–121.
Jin, Y., and J. Fan. 2011. "Test for English Majors (TEM) in China." *Language Testing* 28 (4): 589–596. doi:10.1177/0265532211414852.
Just, M. A., and P. A. Carpenter. 1992. "A Capacity Theory of Comprehension: Individual Differences in Working Memory." *Psychological Review* 99 (1): 122–149. doi:10.1037/0033-295X.99.1.122.
Kane, M. J., D. Z. Hambrick, S. W. Tuholski, O. Wilhelm, T. W. Payne, and R. W. Engle. 2004. "The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning." *Journal of Experimental Psychology: General* 133 (2): 189–217. doi:10.1037/0096-3445.133.2.189.
Ketelsen, K., and M. Welsh. 2010. "Working Memory and Mental Arithmetic: A Case for Dual Central Executive Resources." *Brain and Cognition* 74 (3): 203–209. doi:10.1016/j.bandc.2010.07.011.
Köpke, B., and J.-L. Nespoulous. 2006. "Working Memory Performance in Expert and Novice Interpreters." *Interpreting* 8 (1): 1–23. doi:10.1075/intp.8.1.02kop.
Köpke, B., and T. M. Signorelli. 2012. "Methodological Aspects of Working Memory Assessment in Simultaneous Interpreters." *International Journal of Bilingualism* 16 (2): 183–197. doi:10.1177/1367006911402981.
Kroll, J. F., E. B. Michael, N. Tokowicz, and R. Dufour. 2002. "The Development of Lexical Fluency in a Second Language." *Second Language Research* 18 (2): 137–171. doi:10.1191/0267658302sr201oa.
Kroll, J. F., and E. Stewart. 1994. "Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections between Bilingual Memory Representations." *Journal of Memory and Language* 33: 149–174. doi:10.1006/jmla.1994.1008.
Liu, M., D. L. Schallert, and P. J. Carroll. 2004. "Working Memory and Expertise in Simultaneous Interpreting." *Interpreting* 6 (1): 19–42. doi:10.1075/intp.6.1.04liu.
Mackay, N., A. McCluskey, and R. Mayes. 2010. "The Log Handwriting Program Improved Children's Writing Legibility: A Pretest-Posttest Study." *The American Journal of Occupational Therapy* 64 (1): 30–36. doi:10.5014/ajot.64.1.30.
Mackintosh, J. 1999. "Interpreters are Made Not Born." *Interpreting* 4 (1): 67–80. doi:10.1075/intp.4.1.08mac.
Mizera, G. J. 2006. "Working memory and L2 oral fluency." Doctoral dissertation, the Graduate Faculty of Arts and Sciences, University of Pittsburgh, Pittsburgh.
Moser-Mercer, B. 1985. "Screening Potential Interpreters." *Meta* 30 (1): 97–100. doi:10.7202/003631ar.
Moser-Mercer, B., U. H. Frauenfelder, B. Casado, and A. Künzli. 2000. "Searching to Define Expertise in Interpreting." In *Language Processing and Simultaneous Interpreting: Interdisciplinary Perspectives*, edited by B. Englund Dimitrova and K. Hyltenstam, 107–132. Amsterdam/Philadelphia: John Benjamins.
National Advisory Commission on Foreign Language Teaching in Higher Education. 2000. *Gaodeng Xuexiao Yingyu Zhuanye Yingyu Jiaoxuedagang* [The National College English Teaching Syllabus for English Major]. Shanghai: Shanghai Foreign Language Education Press.
Obler, L. K. 2012. "Conference Interpreting as Extreme Language Use." *International Journal of Bilingualism* 16 (2): 177–182. doi:10.1177/1367006911403199.
Quirmbach, L. M., A. J. Lincoln, M. J. Feinberg-Gizzo, B. R. Ingersoll, and S. M. Andrews. 2009. "Social Stories: Mechanisms of Effectiveness in Increasing Game Play Skills in Children Diagnosed with Autism Spectrum Disorder Using a Pretest Posttest Repeated Measures Randomized Control Group Design." *Journal of Autism and Developmental Disorders* 39: 299–321. doi:10.1007/s10803-008-0628-9.
Russo, M. 2011. "Aptitude Testing over the Years." *Interpreting* 13 (1): 5–30. doi:10.1075/intp.13.1.02rus.

Seleskovitch, D. 1999. “The Teaching of Conference Interpretation in the Course of the Last 50 Years.” *Interpreting* 4 (1): 55–66. doi:10.1075/intp.4.1.07sel.

Service, E., M. Simola, O. Metsaenheimo, and S. Maury. 2002. “Bilingual Working Memory Span Is Affected by Language Skill.” *European Journal of Cognitive Psychology* 14 (3): 383–407. doi:10.1080/09541440143000140.

Shah, P., and A. Miyake. 1996. “The Separability of Working Memory Resources for Spatial Thinking and Language Processing: An Individual Differences Approach.” *Journal of Experimental Psychology: General* 125 (1): 4–27. doi:10.1037/0096-3445.125.1.4.

Turner, M. L., and R. W. Engle. 1989. “Is Working Memory Capacity Task Dependent?” *Journal of Memory and Language* 28 (2): 127–154. doi:10.1016/0749-596X(89)90040-5.

Tzou, Y. -Z., Z. R. Eslami, H. C. Chen, and J. Vaid. 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 (2): 213–227. doi:10.1177/1367006911403197.

Zou, S. 2011. “Tisheng Kaoshi Gongpingxing [On Enhancing Test Fairness].” *Waiyu Ceshi yu Jiaoxue* [Foreign Language Testing and Teaching] 1 (1): 42–50.

Zou, S., and W. Chen. 2010. “Huigu yu Zhanwang [TEM Tests: Past, Present and Future].” *Waiyujie* [Foreign Language World] 30 (6): 9–17.