



Investigating the Declarative-Procedural Gap for the Indirect Speech Construction in L2 Learners

Zhengrong Chen¹ · Catherine Caldwell-Harris²

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

It is common to have good declarative but poor procedural knowledge of a foreign language, especially for classroom learners. To study this gap in a constrained manner, we asked Chinese learners of English to repeat, correct and produce indirect speech. The indirect speech construction was selected in the present study because it is known to be a particularly complex construction. Chinese university students who all had good declarative knowledge of the rules governing indirect speech were selected to have overall low or high oral proficiency when assessed in a free speech situation. High proficiency participants pursued strategies that increased their speech rate while reducing errors. They used more idiomatic English, more chunked expressions, and showed less negative transfer from Mandarin, compared to low proficiency participants. Indeed use of chunks was the primary means by which both groups of participants were able to increase their accuracy, complexity, and speaking rate. Low proficiency but not high participants showed evidence of a speed-accuracy trade-off. They either kept errors low at the cost of high pausing, or produced many errors with the benefit of rapid speech. Identifying preferences for speed versus accuracy could facilitate methods for encouraging learners to move out of their comfort zones.

Keywords The declarative-procedural gap · The indirect speech construction · Oral proficiency · L2 learners

Introduction

A frequent situation for foreign language learners is to master complex grammatical rules after extensive, deliberate training, but then flounder when attempting to use their knowledge in a spoken language task. This is the well-known problem of obtaining fluency in oral production despite good or even excellent written language abilities (see Segalowitz 2010 for a review of issues). This problem is intimately connected to a key topic in the cognitive sciences, which is

✉ Zhengrong Chen
helenczr@126.com

¹ School of Foreign Languages, Southeast University, 2# Southeast University Road, Jiangning District, Nanjing 211189, Jiangsu Province, China

² Department of Psychological and Brain Sciences, Boston University, Boston, MA 02215, USA

understanding the differences and distinct roles of humans' multiple memory systems (Anderson 1995; Cohen and Squire 1980). These include declarative knowledge (knowing-that, also called explicit knowledge; Ullman 2004) and procedural knowledge, as in Ryle's (1949) classic knowing-how view. Declarative knowledge is knowledge of objects and events which is recorded during higher cognitive processes or consciousness and can be accessed by such processes while procedural knowledge is knowledge of how to perform a cognitive skill and is inaccessible to conscious processes (Anderson 1976; Eichenbaum and Cohen 2001; Schacter and Tulving 1994; Squire and Knowlton 2000; Ullman 2004).

It is common to have good declarative but poor procedural knowledge of a foreign language, especially for classroom foreign language learners, but so far few empirical studies of language learning have reported on the investigation of the declarative-procedural gap in oral production. Lack of proceduralization may be a key reason why learners' spoken language may lack fluency and contain errors, even for learners who know the relevant vocabulary items and grammatical rules, and who perform well on written tests. Traditional assessment of L2 oral production tends to be restricted by the influence of Chomsky's and Hymes' ideas on linguistic competence with indices measuring fluency (such as speech rate), accuracy (such as number of pronunciation or lexical choice errors), and complexity (such as vocabulary difficulty or syntactic complexity) (Yang 1999; Zhang and Wu 2001; Zhou 2002). However, according to Chen (2009), linguists and psycholinguists should assess L2 proficiency not only by testing grammatical and vocabulary knowledge, or declarative knowledge but also stressing psychological aspects such as proceduralization and use of formulaic language. Regrettably, less literature has reported the empirical studies of L2 oral proficiency from the angle of proceduralization.

The present study was motivated by the problems in oral production of Chinese tertiary non-English major students in one of the key universities in Nanjing, China, such as disfluency and inaccurate use of language units and grammar rules. Based on the review of literature concerning proceduralization, the present study investigated the declarative-procedural gap by distinguishing the differences in L2 oral production of indirect speech in fluency, accuracy and complexity, probing into speaking strategies and exploring the influence of the task difficulty and chunking differentiating L2 learners with good oral proficiency from those with poor speaking ability in China.

The present study has both methodological and pedagogical significance. Methodologically, it is significant in that it was conducted with a multi-disciplinary and multi-dimensional method, consisting of pretests, three cognitive behavioral tasks and a follow-up interview, which has its advantages over those merely adopting commonly used methods in linguistics. Pedagogically, the study is significant because it may help L2 speaking teachers and students realize the importance of proceduralization in L2 speaking and offer them strategies of how to accumulate and make good use of chunks, idioms and collocations. Extensive practice of L2 speaking has great impact on proceduralization and on L2 learning in reconstructing procedural knowledge, which calls for more attention.

Below the paper will start from the review of literature on proceduralization and then introduce the method, report the results and discuss them.

Proceduralization

Proceduralization is believed to be essential for fluent and appropriate speech because of the high demands that speaking places on working memory (Levelt 1989). Speaking, because of its inherent time demands, requires that a large number of decisions be made in a short period of time (Towell, Hawkins and Bazergui 1996). Speakers must decide what to say (macroplanning), which can require most of one's executive function abilities given the pragmatic and social demands of many speech situations. Once speech content has been decided and communicative goals are in mind, speakers must then perform myriad micro-planning tasks. These include deciding how the ideas are to be communicated, including speech style, point of view, type of words and grammatical structures. The next stages are lexical selection, morphological and grammatical encoding, finally motor and speech planning (Levelt 1989). De Bot (1992) noted that errors and speech dysfluencies can originate at any of these stages, and are especially acute when the speech preparation tasks require piecemeal production using limited-capacity executive function.

Theorists in second language acquisition have increasingly studied the factors that underlie fluency gains by L2 learners. It is found that proceduralization increases fluency and decreases pausing and errors by automating many or even all speech tasks and hence automaticity has been heavily studied by cognitive scientists (Chen 2009; Ma et al. 2017; Pawley and Syder 1983; Towell et al. 1996; Wood 2009).

Two types of procedural language have been discussed in the research literature. The first involves often-used linguistic units such as idioms and collocations. Called formulaic expressions, these expressions can be fluently produced and understood, even if the speaker does not understand the grammar or individual vocabulary words (Chen 2009; Wray 2002). Use of these chunks aids fluent speech because they are stored and accessed as whole chunks, thus avoiding the high attentional demands of retrieving and combining separate words (Miller 1956). The frequency of using chunks correlates with measures of fluency in L2 users (Zhen 2009). If these formulaic expressions were initially learned as whole chunks, minimal proceduralization is required, as long as they can be retrieved and produced as prefabricated units. Another interesting discussion around procedural learning for foreign language learners concerns how formulaic expressions are used. Wray (2012) noted that L2 learners make less use of formulaic expression to aid fluency than might be expected to give the big payoff in terms of speed that such expression confer. She suggested that many formulaic expressions are relatively low in frequency, and serendipity is required to hear them in an appropriate context, when the learner is also prepared to learn and benefit from them.

The second type of procedural language is what can be considered true proceduralization. Learners' explicit, declarative knowledge gradually becomes automated due to extensive practice (Chen 2009). But how does declarative knowledge become automated due to practice? Two routes for this seem plausible, drawing on ideas from automaticity of cognitive skills and motor sequences (e.g., Anderson 1995).

1. *Automatic associations between semantic/conceptual memory and language forms.* When nonverbal concepts (or a nonverbal situation) automatically activate words, phrases or even sentences, one can say that production of this concept has been automated (or proceduralized). Examples that most readers will have experienced are language behaviors or routines, such as saying *thank you* or *I'm sorry*. These can be automatically elicited by a specific situation, such as the social requirement to say *thank*

you or *excuse me*. Automatic activation of words and phrases occurs for most people through-out the day, triggered by appropriate environmental stimuli. As second language learners practice using and thinking in their L2, these environmental stimuli can activate the appropriate L2 words and phrases. What is important for automaticity/proceduralization is that the source activates specific lexical forms automatically, with no (or minimal) conscious effort.

2. *Associations between words*. Once words (or phrases) are mentally activated, words activate other words that co-occur with them either consecutively (such as verbs that frequently follow specific nouns), or because they are semantically associated (e.g., *salt, pepper; black, white*). This has long been observed in studies of word association and sentence production in native speakers (Van Petten 1995). With extensive practice, morphological encoding is automated as the outcome of associated links between words and possible morphological ending; word rules are automated following statistical associations between words and other words that typically follow them; e.g., verb phrases usually follow noun phrases in an SOV (Subject Object Verb) language.

Indeed, our study was initiated by specific problems in proceduralization of declarative knowledge among classroom foreign language learners. Classroom teaching in China has typically emphasized drills on vocabulary and written tests, neglecting use of language in face-to-face situations. Compared with the scores of speaking tests, many university students in China obtain higher scores on written tests such as CET 4 and CET 6 (i.e. Band 4 and Band 6 of College English Test, one of the authoritative national English tests in China). This directed us to study the declarative-procedural gap and provided the rationale for our methodology: to understand spoken dysfluencies and errors that occur even when learners have excellent declarative knowledge of the relevant grammatical constructions.

Methods

Research Questions

Our goal was to investigate the following specific topics.

1. Classroom learners have good or bad spoken output. What are the differences between low and high proficiency participants in L2 oral production of indirect speech in fluency, accuracy and complexity? And what are the strategies or abilities distinguishing those with good versus poor speaking ability?
2. Speech tasks vary along a continuum of difficulty and constraint. How does task difficulty influence indices of oral production for high and low proficiency participants?
3. For classroom foreign language learners, how much does fluency and good accuracy depend on chunking? What individual differences exist in use of chunks?

Research Rationale

The Selection of Oral Production of Indirect Speech

The difficult grammatical construction we selected was the indirect speech construction. Instead of directly quoting speech, as in (a) below, one can indirectly report speech, as in (b)

- (a) He said to Kate, "How is your sister now?"
- (b) He asked Kate how her sister was then.

Transforming direct speech to indirect speech is a difficult task for many English foreign language (EFL) learners and indeed even native speakers can make errors in formal written tasks. The shifts include the change of verbs (*said-asked*), tense change (*is-was*), personal pronoun (*your-her*), word order (*How is your sister* changed to *how her sister was*) and time references (*now-then*).

Mandarin has grammatical constructions for reporting others' speech, but their form is substantially different from those in English. One challenge for EFL learners is that in Mandarin, past tense is indicated by adverbs of time or auxiliary words, instead of changing verb forms as in English. This makes the obligatory tense change in indirect speech difficult. Other difficulties are shifting pronouns for proper names, and changing adverbials of time. Tense, pronouns and time adverbials are among the most difficult aspects of English grammar for EFL learners in China (Li 2008). Mastery of the indirect speech construction is emphasized in many EFL classrooms in China, where students are taught explicit rules for reporting indirect speech, and rigorously tested on their knowledge of it. Educators stress the written grammatical knowledge of this construction only, with little oral practice.

The indirect speech construction is thus a rich construction for understanding limitations on proceduralization of declarative knowledge. However, very little research exists on foreign language learners' written knowledge or oral production of this construction. The current study sought to fill this gap in the research literature. We expected that even Chinese EFL learners with good declarative understanding of indirect speech would have difficulty transforming tense, pronouns and adverbials under the time pressures of spoken production.

Methods for Eliciting Speech Samples

Speech elicitation tasks span a continuum from maximally free response to minimally constrained. Maximally free responses are spontaneous speech samples. Commonly used prompts are to describe yourself or your hobbies, or to prepare a speech on a topic. More constrained methods are picture description, story retelling, and reading (see Segalowitz 2010). Free response tasks are the most demanding because they require both macro and micro planning. Relatively unconstrained elicitation tasks like free description and story retelling will produce dysfluencies that arise from multiple points in the macroplanning and microplanning stages.

Because our interest was in proceduralization of declarative knowledge, we decided to constrain our learners' speech as follows.

- (a) We used three tasks that varied in difficulty, but still constrained content. These were a basic repetition task (repeat the auditory sentence), a judge-and-revise task (repeat the sentence with any grammatical errors fixed) and a report-dialog task (use indirect speech to report what one person said to another). These tasks focus participants on speech production (in the case of the repeat task) and microplanning and speech production in the case of the other two tasks.
- (b) Materials for the tasks involved one grammatical construction, the English indirect speech construction. As described further in the method section, participants were screened for having a high level of declarative understanding of this construction.

Measuring Speech Output

To study the characteristics of proficient oral output, researchers have used both subjective and objective measures.

Subjective Measures

Speech samples can be described using holistic rating criteria. Trained raters rate speech for fluency, which is usually defined as a combination of speaking speed and absence of excessive pausing. Accuracy in grammar, pronunciation, complexity and appropriateness can also be rated. Complexity can reflect syntactic novelty and lexical distinctiveness, such as using a distinctive or slightly rare word rather than a generic, common word. Appropriateness takes into account the pragmatics of the discourse situation.

Objective Measures

Speech can be audio-recorded and subjected to micro-analysis. Well-defined indices exist for measuring fluency and accuracy. These include measuring speech rate (often syllables per minute), length of run (number of syllables without pausing), average length of pause and number of pauses. Pronunciation and grammatical accuracy can also be measured by training raters to compare audio recordings to native speaker outputs.

To identify participants who varied in spoken language ability, we used trained raters to evaluate spontaneous speech samples. We then obtained audio recordings of these participants performing our tasks, and measured fluency using the micro assessments (e.g., pause length, speech rate) that are indices in the literature on spoken language. The auditory language was also assessed for number of chunks used by each participant in each sentence.

Participants

To find a set of English learners who varied in spoken language ability but who all knew the rules for reporting indirect speech, we screened 95 students from one of the key universities in Nanjing, China. These students were administered a grammatical test of indirect speech and a separate test of English speaking abilities (see pretest materials described below), with the result that 32 students were selected for the current study.

The 32 participants (50% female) were non-English majors, aged from 22 to 25, who had studied English as a foreign language for 8–10 years. They had all taken the College English Test (CET), national English as a foreign language test in the People's Republic

of China. This test examines the English proficiency of undergraduate and postgraduate students in China and ensures that students have reached the required English levels specified in the National College English Teaching Syllabus. Participants' scores on the CET 4 ranged from 538 to 610 (highest possible score is 710). This represents above-average ability for Chinese students and indicates a high degree of declarative knowledge of written English grammar and vocabulary.

Pretests

Oral Test

To identify participants with either good or poor spoken English abilities, we administered a test of English speaking abilities to the 95 students. As in the Speaking English Test (SET for short, a subcomponent of the College English Test mentioned above) the oral test consisted of three parts: Greetings, Questions and Answers, and an oral presentation. During the first part, speakers were asked to introduce themselves as much as possible. During Questions and Answers (lasting about 3 min), speakers were asked three yes/no and wh-questions about living in Nanjing. For the oral presentation, speakers addressed the topic "The Advantages and Disadvantages of the Internet" using five words or phrases listed in text box on a printed sheet. They had 1 min to prepare and 2 min for speaking.

Speech data for this oral test was recorded and scored by two experienced raters of the official CET-SET test, using the standard scoring methods of the CET-SET. The scoring standards involve making holistic judgments on syllable fluency and chunk fluency, pausing, sophistication of sentence patterns, pronunciation, syntactic and lexical variety, and information organization. The final score of each item was averaged by the two raters' scores, leading to a range of scores across learners. Based on this scoring, the 36 learners with the highest scores were categorized as being high proficiency in their speaking abilities, and the lowest 48 were categorized as being low proficiency. The 11 test-takers in the middle of the proficiency range were excluded from further consideration.

Grammatical Test of Indirect Speech

To identify participants with perfect explicit knowledge of the English indirect speech construction consisting of 15 multiple choice questions, items were constructed using examples from classroom grammar books. Instructions were in English and no time limit was placed on completion. Of the 75 participants who received full scores on this grammar test, 16 were selected from the high proficiency group, and 16 were selected from the low proficiency groups that had been established using the oral test described above. Selection was at random with the constraint that 8 males and 8 females would be in each proficiency group.

This procedure ensured that participants varied in their spoken English abilities, even though all had good understanding of the grammatical rules for reporting indirect speech. In the remainder of the paper, we will compare fluency and accuracy for learners in the low and high proficiency groups, so it is important to remember that low and high proficiency refers only to spoken language ability, not written knowledge of English vocabulary or grammar.

Experimental Tasks

Three tasks were designed to vary in their processing demands, from relative easy (repeat a correct utterance), to moderate difficulty (orally revise an incorrect sentence), to challenging (listen to a dialogue and then report “what the man said to the woman”).

Repeat Task

Participants listened to audio recordings of 12 sentences containing indirect speech, such as *His mother asked him what he had done the day before*. Participants’ task was simply asked to repeat the sentences they heard. Indirect speech sentences were inter-mixed with 8 filler sentences.

Revise Task

Participants listened to audio recordings of incorrect versions of the sentences described above, along with the same 8 filler sentences (all of which had correct grammar). For each sentence, they were instructed to judge whether the sentence was grammatically correct. If it was grammatically correct, their job was simply to repeat it. If the sentence was grammatically incorrect, they were instructed to produce a grammatically correct version of the sentence.

The twenty sentences used in these tasks contained 8–12 words, a length which could be read within 4 s. Slightly longer sentences used in the pilot study were found to be too long for some participants, and thus were shortened to minimize working memory demands. All audio materials were recorded by an English teacher of Chinese origin who was considered to have excellent English pronunciation. The speed of speaking was approximately 150 words per minute. A pause of 15 s between sentences allowed participants to perform their repeat or judge-and-repeat tasks. Eight grammatical filler sentences were interspersed among those containing indirect speech.

Report Dialog Task

To elicit productive use of English indirect speech forms in a controlled fashion, we asked participants to listen to a dialogue in Chinese, followed by an English question about what one character had said to another. They then had to produce in English answers to questions about which character said something to the other character, as illustrated below.

Dialogue 1:

Man:

小明和小芳還沒有來呢。

Woman:

W: 他們今天上午不可能到達國際展覽中心了。

Question:

What did the woman say to the man?

The dialogue was made between a man and a woman in Chinese as indicated above. Then a question was asked “What did the woman say to the man?” The participants were required to answer the question only in English. We then analyzed whether and how well respondents used English indirect speech forms.

In a pilot study, participants were asked to listen to a Chinese sentence and then interpret it in English. Although we hoped our participants would use the English indirect speech form, they did not. We thus explicitly asked participants to report the Chinese sentences as if they were over-hearing a dialog.

Prior to the tasks, the participants were not informed of our interest in the English of indirect speech form. All tasks were audio-recorded and transcribed for data analysis.

Because the repeat task and the revise tasks used the same items, it was necessary to assign half the participants to one task and half to the other. The report dialog task used different items and thus all 32 selected participants were involved.

Procedure and Data Analysis

A training session was administered to familiarize participants with the repetition task to which they had been assigned. This was followed by a distractor task (solving arithmetic problems), the Report-Dialog Indirect Speech Task, and finally a debriefing interview. All oral productions were transcribed for data analysis.

We obtained measures of fluency, accuracy and complexity for each task.

The Analysis of Fluency

The following five indices were calculated, following prior research (Kormos 2006; Yuan and Guo 2010; Zhang and Wu 2001; Zhou 2002): response duration (seconds), speech rate (syllables per minute), phonation time ratio; average length of run (number of syllables); average length of pause (seconds) (see Table 1).

Chunk Frequency

Chunk frequency was collected and analyzed for exploring the effects of chunks on fluency of oral production (Wei 2007; Yuan and Guo 2010; Zhen 2009). Chunk frequency was the number of chunks per sentence. We followed definitions of chunks in the research literature (Towell et al. 1996; Yuan and Guo 2010). Note that the sentences in our tasks did not use

Table 1 Fluency indices and their calculation

| Indices | Calculation |
|------------------------------------|--|
| Response duration (seconds) | Time from beginning to end of response |
| Speech rate (syllables per minute) | Total number of syllables ÷ total response duration * 60 |
| Phonation time ratio | The time spent speaking (excluding pauses) ÷ total response duration |
| Average length of run (syllables) | Total number of syllables ÷ (total number of pauses + 1) |
| Average length of pause (seconds) | Total time of pausing ÷ total number of pauses |

Pause defined as break of 0.3 s or longer

idioms or formulaic expressions (Wray 2002), but in instead used phrases that are typically used in overseas English instructional materials. Chunks were thus defined as runs of speech with no pauses of 0.3 s or longer which were collocations, fixed expressions, lexical bundles, polywords (such as *by the way*) and sentence builders (such as *I think that*).

The Analysis of Accuracy

Trained raters, blind to the high/low proficiency status of speaker, assessed correctness of pronunciation by listening to the audio-tapes. Correctness at the level of grammar was judged by reading the written transcripts. Pronunciation of words was scored as incorrect if a word could not be identified or if a phoneme was pronounced incorrectly. Pronunciation errors were recorded as number of pronunciation errors per sentence. Raters assessed the following aspects of the grammatical accuracy of the indirect speech form: correctness of sentence subject, predicate verb(s) (tense, voice and word order), objects, conjunction, adverbials of time and adverbials of place. The final index of grammatical errors was the number of errors per sentence.

The Analysis of Complexity in the Report Dialog Task

Following Rauterberg (1993) we measured complexity by asking trained raters to evaluate three aspects of participants' free responses to the prompts: (1) syntactic variety and lexical variety; (2) sentence length; and (3) the degree of difficulty in pronunciation, words, phrases and sentence patterns. These are the same judgments required for scoring the CET-SET, as described earlier, and the raters were experienced at CET-SET scoring. Raters ranked each response on a scale of 1–4, indicating lowest to highest variety, complexity or degree of difficulty.

Examples of degree of difficulty include the comparison between *exhibition hall* and *museum*; the former is more difficult in pronunciation and word choice, by being a less common phrase. For verbs, *purchase* is considered more difficult than *buy*, as it is multisyllabic and includes difficult to pronounce consonant clusters; it is also less frequent. Similarly, *complete* is considered more difficult than *finish* in pronunciation and word choice.

Following CET-SET scoring, sentence length was evaluated by counting the words in the response and comparing it to a target response. If a response for a given dialog was as long as or longer than the target response, that item received the maximal sentence length score of 4. Scores of 3 indicated that the produced sentence length, relative to the target, was fair; 2 indicated adequate, 1 indicated poor.

Post-experiment Interviews

During participants' oral responses of the tasks, the experimenters took notes about unusual aspects that could be investigated as part of a post-experiment, semi-structured interview. Sample topics (in Mandarin, except for the quoted English sentence) included:

- Recall the sentence, *He began to think about what he should do*. Why did you revise this given that it was correct?

- You did a good job when asked to revise the incorrect sentences, but in reporting the dialog, you had a much harder time. What was your opinion about the difficulty of these tasks?”

We also asked participants how they felt about their performance, and about their own experience and practice in using the English indirect speech form and producing oral English in general. Responses were analyzed using qualitative methods (Bernard 1996; Ryan and Bernard 2000). Although it is beyond the scope of the current paper to analyze interviews in full, we will refer to them for some examples and insights into participants' responses.

Results and Discussion

We first review how fluency and accuracy measures varied across the three tasks for the low and high proficiency participants. Then we present results for complexity measures and chunking, and discuss individual differences in speaking, strategies applied by high proficiency participants and teaching implications.

Results

Fluency and Accuracy

On average, the high proficiency participants were more fluent and more accurate on all indices (see Fig. 1 and Tables 2, 3). As expected, the simple repeat task was the easiest and led to overall low errors and good fluency for both proficiency groups. Compared to the repeat task, accuracy and fluency were poorer for the revise task and were even worse (for most measures) for the report indirect speech task. This is not surprising given that this task required the most planning of which lexical items to use.

The different tasks yielded evidence of qualitative differences between the groups (see Fig. 1). Average pause length has been plotted as a representative measure of fluency. Both pronunciation and grammatical accuracy were plotted in Fig. 1 as they led to different patterns between the proficiency groups.

The report dialog task was the most difficult task for the all participants for most measures. Unexpectedly, the high proficiency participants had similar grammatical errors in the revise and report dialog tasks. In contrast, the grammatical errors of low proficiency participants soared in the report dialog task, relative to the revise task. Note that the report dialog task required more microplanning. The revise task required detecting and orally fixing a single grammatical error. Detecting grammatical errors is well within the competence of all participants given their good scores on the indirect speech grammar task. The report dialogue task required several grammatical elements to be transformed (see Dialogue 1). It also required participants to make choices about how much paraphrasing to do compared to using the exact words in the dialog. The fact that the high proficiency participants had a similar number of errors in the revise task and the report dialog task thus needs to be explained. For the measures of average pause length and grammatical errors, the low proficiency participants can also be seen to have special problems with the report dialog task.

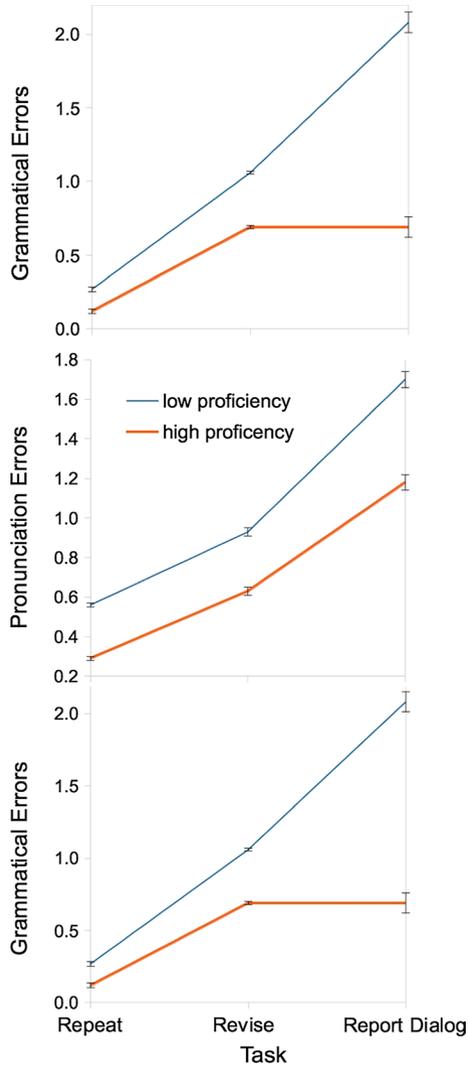


Fig. 1 How fluency and accuracy vary by tasks for high and low proficiency speakers

Scrutiny of the written transcripts clarified differences in strategies and styles of the two groups; these illuminate the group differences just discussed. The low proficiency participants were overall more rigid in their English productions. In particular, L1 interference appeared to have caused many of the low proficiency participants' errors in vocabulary and grammar (shown in Table 3).

Below are examples showing participants being influenced by the differences between Mandarin and English. Participants were confused about the placement of auxiliary verbs, question words and time adverbials. In Mandarin, auxiliary verbs are not used to indicate future time, creating uncertainty for learners about placement of the English auxiliary verb. In the object clause in English, the adverbial of time must

Table 2 Comparison of indices of fluency for three tasks

| | High-proficiency participants | | Low-proficiency participants | | Mean difference | t | Sig. <i>p</i> value |
|---|-------------------------------|------|------------------------------|------|-----------------|------|---------------------|
| | Mean | SD | Mean | SS | | | |
| Repeat indirect speech task | | | | | | | |
| Response duration (seconds) | 4.0 | 0.1 | 4.9 | 4.0 | -0.9 | 14.6 | 0.000 |
| Speech rate (syllables per minute) | 194 | 4.4 | 159 | 194 | 35 | 12.6 | 0.000 |
| Phonation time ratio | 0.96 | 0.02 | 0.81 | 0.96 | 0.15 | 17.3 | 0.000 |
| Average length of run (number of syllables) | 12 | 0.4 | 9 | 12 | 3 | 20.9 | 0.000 |
| Average length of pause (seconds) | 0.4 | 0.04 | 0.6 | 0.4 | -0.2 | 12.2 | 0.000 |
| Revise indirect speech task | | | | | | | |
| Response duration (seconds) | 4.4 | 0.07 | 5.1 | 0.43 | -0.7 | 4.1 | 0.004 |
| Speech rate (syllables per min) | 165 | 8.1 | 139 | 15.2 | 26 | 4.2 | 0.001 |
| Phonation time ratio | 0.81 | 0.01 | 0.70 | 0.07 | 0.11 | 4.1 | 0.004 |
| Average length of run | 12 | 0.5 | 8 | 1.1 | 3.33 | 8.1 | 0.000 |
| Average length of pause | 0.6 | 0.02 | 0.7 | 0.1 | -0.1 | 5.0 | 0.001 |
| Report-dialog indirect speech task | | | | | | | |
| Response duration (sec) | 6.1 | 0.12 | 7.5 | 0.50 | -1.4 | 10.8 | 0.000 |
| Speech Rate (syllables per min) | 170 | 9.6 | 134 | 9.1 | 36 | 10.9 | 0.000 |
| Phonation time ratio | 0.8 | 0.01 | 0.6 | 0.07 | 0.2 | 11.9 | 0.000 |
| Average length of run | 11 | 0.1 | 8 | 0.4 | 3 | 30.9 | 0.000 |
| Average length of pause | 0.9 | 0.07 | 1.5 | 0.07 | -0.6 | 26.8 | 0.000 |

Table 3 Comparison of indices of accuracy for three tasks

| | High-proficiency participants | | Low-proficiency participants | | Mean difference | t | Sig. <i>p</i> value |
|--|-------------------------------|------|------------------------------|------|-----------------|------|---------------------|
| | Mean | SD | Mean | SD | | | |
| Repeat task | | | | | | | |
| Mispronounced words per sentence | 0.29 | 0.04 | 0.56 | 0.06 | -0.27 | 10.5 | 0.000 |
| Grammatical errors per sentence | 0.12 | 0.06 | 0.27 | 0.04 | -0.15 | 5.9 | 0.000 |
| Revise incorrect indirect speech task | | | | | | | |
| Mispronounced words per sentence | 0.63 | 0.10 | 0.93 | 0.13 | -0.30 | 5.2 | 0.000 |
| Grammatical errors per sentence | 0.69 | 0.04 | 1.06 | 0.10 | -0.37 | 10.1 | 0.000 |
| Correct sentences (out of 12) and % | 7.75 64% | 0.89 | 5.13 42% | 1.13 | 2.62 22% | 5.2 | 0.000 |
| Report dialog task | | | | | | | |
| Mispronounced words per sentence | 1.18 | 0.25 | 1.70 | 0.16 | -0.52 | 7.0 | 0.000 |
| Grammatical errors per sentence | 0.69 | 0.26 | 2.08 | 0.18 | -1.39 | 17.7 | 0.000 |
| Correct sentences (out of 6) and % | 3.69 61% | 0.60 | 2.00 33% | 0.52 | 1.69 28% | 8.5 | 0.000 |

Correlations all significant at 0.01 level (2-tailed). Number of correct sentences and measures of complexity only pertinent to the report dialog task

appear at the end of the sentence. In Mandarin it may appear after the subject. In the interrogative sentence in English, the question word (e.g. *when* or *where*) should be put at the beginning of the object clause, while in Chinese it may appear in the middle of the sentence.

Dialogue 2:

| | |
|-----------|--|
| Man: | 這本書很有趣。(Man: The book is very interesting.) |
| Woman: | 湯姆,我明天在什麼地方能買到這本書啊?(Woman: Tom, where will I buy the book tomorrow?) |
| Question: | What did the woman ask the man? |
| Correct: | The woman asked the man where she would buy that book the next day. |

Incorrect responses showing the influence of Chinese word order:

1. The woman asked the man she where would buy that book the next day.
2. The woman asked the man she tomorrow where would buy that book.
3. The woman asked the man where would she buy that book the next day.

It was on the report dialog task where low proficiency participants were most likely to err by using Mandarin word order. High proficiency participants rarely produced sentences with errors resulting from using Chinese word order. This likely reduced their overall errors.

Complexity

Utterances in the report dialog task of high proficiency participants had greater lexical and syntactic variety than those of the low proficiency participants (see Table 4). As an example, consider Dialogue 1. Analysis of the scripts of the participants' oral production revealed that most of the high proficiency participants translated “國際展覽中心” into English as *the International Exhibition Center* while most low proficiency participants preferred to use *the International Show Center*. The word *exhibition* is morphologically more complex than the word *show*. In the post-experiment interviews, we asked participants about the reason for choosing *show* even though the technically correct translation was *exhibition*. Low proficiency participants admitted that the word *exhibition* was hard to pronounce and so they substituted *show*.

Given that low proficiency participants sometimes avoided complex words, it is a bit surprising that average differences between low and high proficiency participants on the

Table 4 Complexity in report dialog task

| Items | High-proficiency participants | | Low-proficiency participants | | Mean difference | t | Sig. <i>p</i> value |
|-------------------------------|-------------------------------|------|------------------------------|------|-----------------|-----|---------------------|
| | Mean | SD | Mean | SD | | | |
| Syntactic and lexical variety | 2.70 | 0.25 | 2.16 | 0.21 | 0.54 | 6.6 | 0.000 |
| Length of sentence | 2.72 | 0.25 | 2.21 | 0.14 | 0.51 | 7.0 | 0.000 |
| Degree of difficulties | 2.69 | 0.24 | 2.18 | 0.18 | 0.51 | 6.8 | 0.000 |

three measures of complexity were not large. While significant, t values are smaller than many of those obtained for fluency and accuracy differences. One reason for this is that low proficiency participants sometimes produced more complex sentences than did high proficiency participants. We can explain this as a byproduct of low proficiency participants' tendency to translate from Chinese. This increased their grammatical errors (as noted in the discussion of accuracy above), but also sometimes increased complexity. The reason is that translations from Chinese, when correct, were sometimes grammatically more complex than idiomatic English.

An example of greater use of idiomatic English occurred when answering the question in the third dialogue “你騎自行車還是乘公車上學啊?” Many of the high proficiency participants produced a fairly simple sentence, *The woman asked the man whether/if he went to school by bike or by bus.* In contrast, four low proficiency participants produced the sentence, “The woman asked the man whether/if he rided/rode his bike (bicycle) to go to school or took bus/buses to go to school.” This is a complex syntactic production because of the two verb forms, which both require a tense change due to the rules of reporting of indirect speech. By using the simpler sentence, the high proficiency participants only needed to change the tense of one verb, *go*. The high proficiency participants may have been more familiar with lexical bundles such as *go to school* and *by bike* or *by bus*, and thus were more fluent and flexible in oral production. The presence of these cases suggests the high proficiency participants were more flexible language users, using reduction strategies to produce a simpler output, and avoidance strategies, to avoid using two verbs (Færch and Kasper 1983).

Chunking

A central question for procedural language is how chunking influences fluency, accuracy and complexity. Table 5 shows correlation coefficients for these indices with chunk frequency (number of chunks per sentence), across the 32 participants. Because the repeat and revise tasks used all of the 16 participants, these tasks were included together when calculating correlation coefficients. This was done after preliminary analysis showed that the correlations obtained when tasks were separately analyzed were similar to each other.

Chunk frequency was strongly correlated with all fluency measures (correlations ranging from $r=0.72$ to $r=0.98$) and with the three measures of complexity (correlations ranging from $r=0.89$ to $r=0.94$). Chunk frequency was moderately to strongly correlated with accuracy, with correlations ranging from $r=-0.53$ (for grammatical accuracy in the repeat + revise tasks) to r values ranging from $r=0.75$ to $r=-0.86$ in the other tasks and measures. This is strong support for the contention that fluent, accurate, complex output is facilitated by chunking.

Two correlations are plotted in Figs. 2 and 3 to show the degree of separation between high and low proficiency participants. Figure 2 plots the average mean length of run (number of syllables produced without pausing) as a function of number of chunks per sentence, for the repeat + revise tasks. As explained in the methods section, chunks were defined as a “run” (meaning a speech sequence without a pause), with the additional requirement that the string of words was a lexical bundle, sentence builder, etc. We infer that the tight clustering of data points around the trend line (with $r=0.98$) occurred because in the repeat + revise tasks, speaking without pausing was usually only possible when producing a chunk or a sequence of chunks.

Table 5 Average correlations between chunk frequency and the fluency, accuracy and complexity

| | Response duration | Speech rate | Phonation time ratio | Length of run (syllables without pausing) | Length of Pause | Grammatical errors | Pronunciation errors | Number of correct sentences | Syntactic and lexical variety | Length of sentence | Degree of difficulties |
|-----------------------|-------------------|-------------|----------------------|---|-----------------|--------------------|----------------------|-----------------------------|-------------------------------|--------------------|------------------------|
| Repeat + revise tasks | -0.91 | 0.86 | 0.81 | 0.98 | -0.87 | -0.53 | -0.75 | | | | |
| Report-dialog task | -0.70 | 0.77 | 0.71 | 0.72 | -0.73 | -0.84 | -0.86 | 0.78 | 0.92 | 0.89 | 0.94 |

Correlations all significant at 0.01 level (2-tailed). Number of correct sentences and measures of complexity only pertinent to the report-dialog task

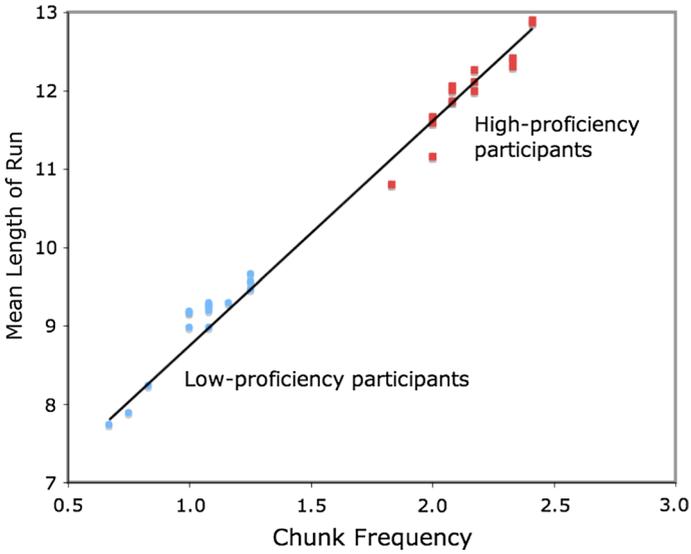


Fig. 2 Average number of syllables without pausing (mean length of run) as a function of chunk frequency in the repeat and revise tasks

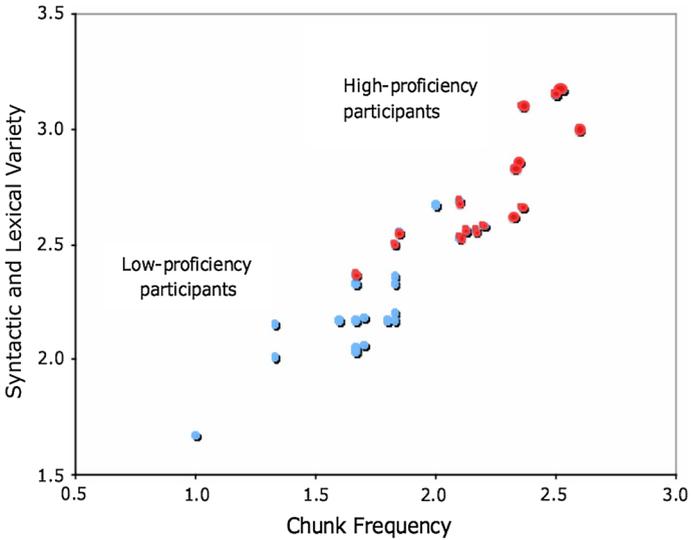


Fig. 3 Syntactic and lexical variety as a function of chunk frequency in the report dialog task

The high proficiency participants typically combined together 2 or 3 chunks to have, on average, a long run of about 12 syllables. This can amount to all or most of a whole sentence (e.g., *The woman suggested the man go see the doctor* is 13 syllables). In contrast, the low proficiency participants used on average only one chunk per sentence, and

thus were able to produce only 8–9 syllables without pausing. Although all high proficiency participants had consistently longer runs and used more chunks than all low proficiency participants, these two groups showed the same strength of correlation between runs and chunks. We infer that using chunks confers similar fluency benefits on both low and high proficiency participants.

Figure 3 plots average ratings of lexical and syntactic variety as a function of chunk frequency in the report dialog task. Participants whose utterances were rated as having less variety than 2 chunks per utterance were also mostly low proficiency.

The high proficiency participants also produced reported dialog that was judged to have greater variety, and they also used more than 2 chunks per sentence. As in Fig. 2, the correlation between lexical and syntactic variety and chunking ($r=0.94$) was very strong, and r values were the same for both high and low proficiency participants (both $r_s=0.83$, note that correlation for each group is lower than for all participants together because of the restricted range of scores resulting from conducting correlations on separate groups).

Individual Differences in Fluency and Accuracy

As discussed earlier, the high proficiency participants performed better than low proficiency participants on all indices of speech production that were measured. An implication is that on the path to higher proficiency, learners tend to reduce their error rate and increase their fluency. However, for these participants, a plot of pronunciation accuracy as a function of speech rate in the report dialog task belies this simple idea (see Fig. 4). The overall correlation, holding across the two groups, between accuracy and fluency in Fig. 4 is impressive, $r=-0.73$. However, as Fig. 4 shows, accuracy and fluency were uncorrelated within proficiency groups. The low proficiency participants, occupying the upper left quadrant of the Fig. 4, did not fall along a trend line of decreasing error rate as a function of increasing speech rate. Indeed, when separate correlations were run on the two groups,

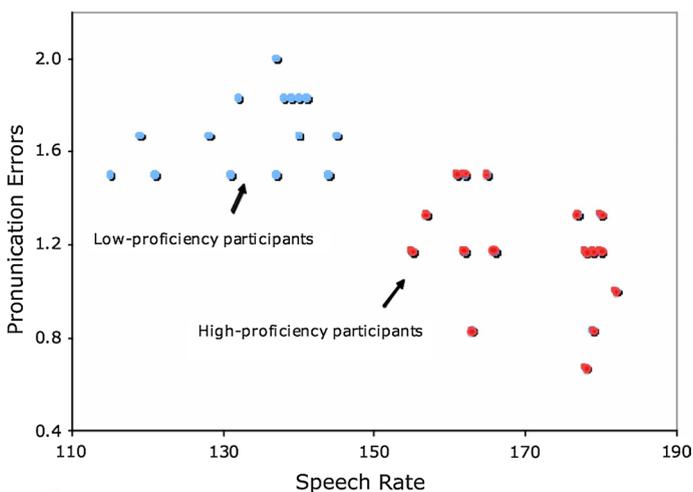


Fig. 4 Pronunciation errors as a function of speech rate reveals speed-accuracy trade-offs for low proficiency bilinguals in the report dialog task

low proficiency participants actually had a slightly positive correlation between errors and fluency $r=0.40$ (although this was not statistically significant due to only 16 participants). This positive correlation means that higher fluency was associated with more errors, suggesting a speed-accuracy trade-off. Speed-accuracy trade-off is the concept that actors can generate behavior more rapidly if they tolerate a higher probability of error, or they can reduce error by reducing speed (Mackay 1982). Note that high proficiency participants had less evidence of a speed-accuracy tradeoff, since they had $r=-0.38$ (also n.s.).

Discussion

Within the literature on second language acquisition (SLA), there has been a recent plea by Segalowitz (2010) for a more thorough understanding of factors that aid fluent speech of second language learners. The current paper responds to this plea by studying characteristics of oral output among English foreign language (EFL) learners residing in Nanjing, China.

Our goal was to study a grammatical construction where students would have explicit knowledge of the grammatical rules, but imperfect ability to produce the construction on demand in a spoken production task, where procedural abilities would be important. The grammatical construction of indirect speech accomplished this. Even in our relatively constraining tasks, low proficiency participants had, on average, one pronunciation error and one grammatical error per sentence.

Characteristics of Low and High Proficiency Participants

High proficiency participants were quantitatively better than the low proficiency participants on all measures, but their superior abilities were most apparent in the easiest task (the simple repeat task) and on the hardest task (the report dialog task). We suggest this is because the medium-difficulty task, the revise task, was the task where having excellent declarative knowledge was the crucial factor for minimizing errors and dysfluencies; it is precisely in declarative knowledge that the groups were most similar. This is a task that most resembles classroom exercises and written drills. When the good grammatical knowledge of the low proficiency participants was either less necessary for the task (as in the simple repeat task) or insufficient (as in the report dialog task, where rapid choices has to be coordinated), the low proficiency participants had substantially reduced performance compared to the high proficiency participants.

High proficiency participants were qualitatively different from low proficiency participants in several ways. They used more idiomatic English and showed less negative transfer from Mandarin grammar than did low proficiency participants. The complexity and lexical/syntactic variety of their sentences were greater, but the high proficiency participants also more frequently produced idiomatic English. This meant they sometimes used simple grammatical construction, leading to a low complexity score for some sentences.

Reduced lexical variety for low proficiency participants was sometimes a result of their desire to choose a simpler word because it would be easier to pronounce. This is consistent with De Bot's (1992) observation that L2 speakers are often strategic in formulating the preverbal message. L2 speakers may know what they want to say, but not know all of the lexical items or correct grammatical structures. In these cases, speakers can strategically

formulate the preverbal message to avoid their linguistic limitation. But these strategies frequently negatively impact fluency, by slowing down formulation of the preverbal message. Thus, one reason for lengthy pausing shown by the low proficiency participants in the report dialog task (see Fig. 1) is likely to have been the need for strategic planning because of the pronunciation difficulties that accompany complex words.

Role of Chunks

High proficiency participants used substantially more chunks than did low proficiency participants (see Figs. 2, 3). Chunk frequency had very strong correlations with fluency. For these tasks and this group of participants, it seemed that use of chunks was the primary means by which participants were able to produce a sequence of syllables without a 0.3 s pause. This is consistent with prior work showing the processing advantages of using chunks (Tremblay et al. 2011). But chunk correlations also had high correlations with accuracy and complexity. Use of chunks appears to be a primary means by which this group of foreign language learners increased their accuracy and complexity.

It is worth noting that chunking may not be causing the observed lexical and syntactic variety. Instead, both variety and frequent chunking may reflect individuals' broader oral language abilities. That is, those EFL learners who have good knowledge of chunks are also those who use more varied syntax and vocabulary. Content analysis of the chunks could be pursued in future work to determine if chunks are more varied than non-chunks, as they plausibly could be, since idioms, and polywords are often syntactically irregular (Wray 2002).

A major question in literature on fluency is how foreign language learners proceed from first knowing declarative rules to fluently enacting them (see Segalowitz 2010). We established that all three aspects of proficient oral output (fluency, accuracy and complexity) are highly correlated with use of formulaic language. This raises the possibility that oral proficiency gains are mainly due to increasing use of chunks, not to proceduralization of formerly declarative knowledge. With learning and practice, learners are exposed to chunks and learn these separately from their declarative knowledge. On this view, declarative knowledge of grammatical rules is available to support untimed, written language, and slow, effortful spoken language, but only automatized chunks allows rapid accurate speech, with little pausing. Future work can investigate the veracity of this proposal.

Speed-Accuracy Trade-Offs

Figure 4 reveals a cluster of a few participants with the highest errors for their group and medium fluency. These participants thus attained a higher speech rate at the expense of poorer pronunciation. Another cluster of 3 participants had the lowest pronunciation error rate, but also had the lowest speech rate; these learners privileged accuracy over fluency in their oral production. What is of particular interest here is that the low proficiency participants showed the speed-accuracy trade-off, not the high proficiency participants. We infer that one aspect of having low proficiency in oral production is that learners make decisions about what components of oral production are most important to them, and they place their energies in that direction.

It will be important to investigate what cognitive and personality styles co-occur with favoring error reduction at the expense of speed or vice versa. For example, low working memory may put a burden on lexical retrieval and morphological encoding, such that persons with low working memory may opt to prioritize error reduction. Learners who are extroverted may care more about maintaining a socially-appropriate speed of conversation, and thus may care less about making errors. If cognitive and personality styles influences speed-accuracy trade-offs, lessons or tutoring can be tailored to be more appropriate for individuals learners.

Teaching Implications

This study revealed a wide gap in ability: perfect scores on a grammar test co-occurred with substantial errors and dysfluencies when speaking even in highly constraining tasks. In China, and in many parts of the world, instruction is focused on achieving good scores on written tests. This produces good declarative understanding but poor proceduralization. Steps need to be taken in the classroom to develop a better balance between written and spoken competence.

We identified two types of strategies that have clear implications for teaching. Learners with poor spoken proficiency were sometimes strategic, choosing an easy-to-pronounce noun (e.g., *show*) in order to avoid the more complex but appropriate word (*exhibition*). Teachers can explicitly discuss this type of strategy, to help students strategize wisely. The speed-accuracy trade-offs we observed can be discussed in class, allowing students to be aware of their particular style. Role-play of dramatic situations can be designed to encourage students to consciously step out of their comfort zone. For example, students could pretend they are sports announcers, speaking play-by-play for the audience (as in transcripts of announcers in Kuiper 1996).

We found strong relationships between frequency of using chunks and all of the measures of proceduralization (fluency, accuracy, and complexity). An obvious implication for teachers is to discuss the benefits of chunking for aiding oral output and to devote class time to practicing use of chunks. Indeed, detailed descriptions of classroom syllabi and methods for using formulaic language have been available for several decades (Nattinger and DeCarrico 1992; Lewis 1997; Wood 2002).

Reasons exist to be cautious before jumping on the formulaic language bandwagon. Wray (2012) noted that many formulaic expressions used by native speakers are quite rare, and may only be encountered in specific situations. Swan cautioned teachers, “Chunks in the classroom: Let’s not go overboard” (Swan 2006, p. 5). His article noted that if learners practiced 10 expressions a day, it would take them 30 years to master the 100,000 fixed expressions in native speakers’ repertoires. He thus advocated focusing on the most common expressions, and urging students to themselves be alert to the expressions that are most useful for their area of specialization. Our work suggests some qualifications to these useful proposals. Swan’s 100,000 expressions refer to non-compositional or only mildly compositional expressions such as *break even* or *out of work*. Our sentences were mostly free of formulaic expressions. Our chunks were collocations (*the day before*, *a cup of tea*) and word sequences that fit together semantically and pragmatically, such as *go to see the film*, *go see the doctor* and *let’s go dancing*. Students can be encouraged to identify these chunks in their English reading material.

Finally, follow-up interviews indicated that besides the chunks learned in textbooks and English class, some of the participants acquired chunks through watching English videos,

listening to English songs including pop music where lyrics are salient, like rap. Future research can investigate how classroom learners are able to incorporate chunks learned from media sources into their oral speech, either via classroom exercises (such as the tasks in the current study) or unscripted social language.

Conclusions and Future Research

As a whole, there exist significant differences between high-proficiency and low-proficiency English speakers in fluency, accuracy and complexity in indirect speech. Under the equivalent condition of the grasp of declarative knowledge, the indices in oral production are related to the application of procedural knowledge or pre-compiled knowledge. The major findings are presented as follows.

1. In general, when compared between levels and between task groups, the high proficiency participants performed better in fluency, accuracy and complexity than the low proficiency peers in oral production of the object clause structure in indirect speech. To further investigate the strategies of high proficiency participants, on the basis of the follow-up interview, we found that the major reasons for this phenomenon were because they were better at the retrieval and use of procedural knowledge.
2. Task difficulty was found to influence the mean differences between the two groups of participants. The major reasons may include the followings. First of all, the high proficiency participants might be more familiar with such difficult structures due to more exposure to the object clause structure in indirect speech from novels, movies, videos and other approaches. Thus the finding is also consistent with schema theory which points out that schemata or mental representations can influence the participants' comprehension and memory.
3. The low proficiency participants might sacrifice their accuracy in order to speak fast, which is consistent with speed-accuracy trade-off relation. According to the follow-up interview, it was found that such phenomenon was because in such case, these participants overstressed fluency while ignoring accuracy in their oral production. This finding implies that in the training of L2 speakers, L2 teachers should use different teaching materials for learners with different proficiency.
4. Compared with the low proficiency English speakers, the high proficiency English speakers more frequently applied the chunks in their use of the object clause structure in indirect speech. Our study found that chunk frequency had positive correlation with fluency, accuracy and complexity, which was proved by the follow-up interviews made with the participants. The finding implies that L2 learners could be encouraged to fall back on the proper use of chunks and other formulaic terms to improve their fluency, accuracy and complexity in oral production.

These findings highlight how many aspects of the effects of procedural knowledge/proceduralization on L2 oral proficiency remain to be empirically explored and confirmed.

Our study still has some limitations in the limited sample size and lack of longitudinal studies. In the future, the follow-up studies will be implemented by enlarging the sample size and adding longitudinal studies to investigate more issues in L2 speaking. Besides, linguists and researchers of oral production could investigate issues such as the following.

1. Grasp of chunks

The current study identified the positive correlation between chunk frequency and fluency, accuracy and complex in oral production. Linguists and researchers may be interested in the way of grasping more chunks in EFL learners' English learning. In such case, teachers might design activities in and out of class to help their students easily access to formulaic language.

2. Personality basis of speed-accuracy trade-off

The current study found out that speed-accuracy trade-off phenomenon existed among the participants. Some of the participants cared more about fluency while sacrificing accuracy while some others paid more attention to accuracy which decreased their speech rate. It might be interesting to have further research on the individual differences such as personality issues of participants. For example, more questions could arise like: Are they extroverts or introverts? If personality influences the performance of the participants, we may design some personality-oriented programs to help such participants to improve their English proficiency and avoid speed-accuracy trade-off phenomenon.

3. Approaches to recognize the effects of proceduralization

It would be possible to conduct longitudinal studies by training some participants to improve their oral proficiency with the principles of procedural memory. For example, researchers may design training plans based on the principles of proceduralization.

Acknowledgements This study was supported by Jiangsu Province Social Science Research Foundation (Grant No. 18YYB008) and Educational Science Foundation during the 13th Five-Year Plan Period in Jiangsu Province (Grant No. B-b/2018/01/45). Additionally, we would like to thank all the participants involved in the study. Finally, our special thanks would be given to the reviewers for their constructive feedback on our previous manuscript and to the editors for their efforts and professional suggestions.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Human and Animal Rights The authors declare that all procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee. This article does not contain any studies with animals performed by any of the authors.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

- Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1995). *Learning and memory: An integrated approach*. New York: Wiley.
- Bernard, H. R. (1996). Qualitative data, quantitative analysis. *Cultural Anthropology methods Journal*, 8(1), 9–11.
- Chen, K. S. (2009). Multiple memory system and foreign language learning. *Foreign Languages Research*, 2, 63–66.

- Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that. *Science*, *210*, 207–210.
- de Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics*, *13*, 1–24.
- Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Memory systems of the brain*. New York: Oxford University Press.
- Færch, C., & Kasper, G. (1983). *Strategies in interlanguage communication*. London: Longman Group Limited.
- Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Kuiper, K. (1996). *Smooth talkers: The linguistic performance of auctioneers and sportscasters*. Mahwah, NJ: Erlbaum.
- Levelt, W. J. M. (1989). *Speaking*. Cambridge, MA: The MIT Press.
- Lewis, M. (1997). *Implementing the lexical approach: Putting theory to practice*. Hove, England: Language Teaching Publications.
- Li, S. J. (2008). Teaching practice of direct speech and indirect speech. *Testing (Edition of Teaching Research)*, *11*, 66–67.
- Ma, D. M., Yu, X. R., & Zhang, H. M. (2017). Word-level and sentence-level automaticity in English as a foreign language (EFL) learners: A comparative study. *Journal of Psycholinguistic Research*, *46*, 1471–1483.
- MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review*, *89*, 483–506.
- Miller, G. (1956). The magical number seven, plus or minus two. *Psychological Review*, *63*, 81–97.
- Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: native like selection and native like fluency. In J. C. Richards & R. Schmidt (Eds.), *Language and communication*. London: Longman.
- Rauterberg, M. (1993). AMME: An automatic mental model evaluation to analyze user behavior traced in a finite, discrete state space. *Ergonomics*, *36*(11), 1369–1380.
- Ryan, G. W., & Bernard, H. R. (2000). Data management and analysis methods. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 769–802). Thousand Oaks, CA: Sage Publications.
- Ryle, G. (1949). *The concept of mind*. London: Hutchinson.
- Schacter, D. L., & Tulving, E. (Eds.). (1994). *Memory systems*. Cambridge, MA: The MIT Press.
- Segalowitz, N. (2010). *The cognitive bases of second language fluency*. New York: Routledge.
- Squire, L. R., & Knowlton, B. J. (2000). The medial temporal lobe, the hippocampus, and the memory systems of the brain. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed.). Cambridge, MA: The MIT Press.
- Swan, M. (2006). Chunks in the classroom: Let's not go overboard. *Teacher Trainer*, *20*, 5–6.
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, *17*(1), 84–119.
- Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, *61*, 569–613.
- Ullman, M. T. (2004). Contributions of neural memory circuits to language: The declarative/procedural model. *Cognition*, *92*, 231–270.
- Van Petten, C. (1995). Words and sentences: Event-related brain potential measures. *Psychophysiology*, *32*, 511–525.
- Wei, N. X. (2007). Phraseological characteristics of Chinese learners spoken English: Evidence of lexical chunks from COLSEC. *Modern Foreign Languages*, *8*, 280–291.
- Wood, D. (2002). Formulaic language in acquisition and production: Implications for teaching. *TESL Canada Journal*, *20*, 1–5.
- Wood, D. (2009). Effects of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study. *Canadian Journal of Applied Linguistics*, *1*, 39–57.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics*, *32*, 231–254.
- Yang, H. Z. (1999). Principles of designing College English Speaking Tests. *Foreign Language World*, *3*, 48–57.
- Yuan, P., & Guo, F. R. (2010). Formulaic language and L2 oral fluency. *Foreign Language World*, *1*, 54–62.
- Zhang, W. Z., & Wu, X. D. (2001). Development of L2 oral fluency: A quantitative analysis. *Modern Foreign Languages*, *24*(4), 341–351.

- Zhen, F. C. (2009). Correlation between chunks, fluency, accuracy and appropriateness of oral English: An empirical study based on COLSEC corpus. *Foreign Language Education in China*, 11, 14–24.
- Zhou, A. J. (2002). Fluency and accuracy improvement in oral English. *Foreign Language Education*, 5, 72–83.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.