International Journal of Applied Linguistics & English Literature
E-ISSN: 2200-3452 & P-ISSN: 2200-3592
www.ijalel.aiac.org.au

Task Complexity and Grammatical Development in English as a Second Language

Satomi Kawaguchi*, Yuan Ma

School of Humanities and Communication Arts, Bilingual Research Lab, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia

Corresponding Author: Satomi Kawaguchi, E-mail: s.kawaguchi@westernsydney.edu.au

ARTICLE INFO

Article history
Received: September 23, 2018
Accepted: December 24, 2018
Published: January 31, 2019
Volume: 8 Issue: 1
Advance access: December 2018

Conflicts of interest: None
Funding: None

ABSTRACT

This study investigates whether second language learners' interlanguage system changes according to the tasks one performs. This has important implications concerning the validity of language testing and assessment. In particular, this study tests Pinemann's (1998) Steadiness Hypothesis within Processability Theory (PT), which states that the basic nature of the interlanguage system does not change across different communicative tasks provided they are testing the same skill type. Our informants are 30 Chinese L1 learners who are learning English as a second language (L2), 10 for each of the three IELTS bands: Low-Intermediate (IELTS 4.5-5.0), Intermediate (IELTS 5.5-6.5) and High (IELTS 7.0 or above). The informants performed two oral production tasks involving different cognitive complexities, which manipulated the variable of +/- planning time based on Robinson's (2011) Cognitive Hypothesis. Analyses revealed that the within-learner competence measured by Pienemann's L2 developmental stages was quite stable in terms of syntactic complexity regardless of the cognitive complexity of the task. On the other hand, L2 learners' performances, measured by rule applications and grammatical accuracy, varied especially with lower level learners. Thus the study investigates whether L2 competence as defined by PT varies according to task complexity variables as defined by Cognitive Hypothesis. The results bridge a gap between two unrelated theories.

Key words: Steadiness Hypothesis, Processability Theory, Task Complexity, Cognitive Hypothesis, English as a Second Language (ESL)

INTRODUCTION

This study aims to investigate how second language (L2) learners' interlanguage (IL) system changes according to the tasks one performs. It is a long-debated issue in the field of L2 acquisition and in language learning pedagogy, as it may have implications concerning the validity of language testing and assessment. That is, if a learner's IL system varies according to the task intended, which task should we use for language assessment? Pienemann's Steadiness Hypothesis (1998) states that the basic nature of the IL system does not change across different communicative tasks provided they are testing the same skill type. Pienemann claims that it is the learner's L2 developmental stages, but not the nature of different tasks, which influence linguistic performance. Other scholars (Tarone, 1985; Bayley & Tarone, 2012) claim that IL is systematically and predictably variable across tasks. However, whether this variability refers to accuracy or to acquired knowledge is not clear. One further important element missing in the debate is the definition of "different tasks". More recently, however, Robinson's Cognition Hypothesis (2003; 2011) provides clear criteria for classifying tasks in terms of cognitive complexity, task conditions and task difficulty. This gives us opportunities to investigate the issue on relationship between IL system and different tasks by utilising Robinson's cognitive complexity of tasks. Thus, the research question in this study is: *Do learners' IL systems vary according to tasks of different degrees of cognitive complexity?* A goal of this study is to test two competing positions of task performance in second language, ie., Pienemann's (1998) steadiness position and Tarone's (e.g., 1985) variability position. No previous studies have tested variability/steadiness of IL systems using tasks of different cognitive complexity. Thus a novelty of our study is to bridge two unrelated theories in SLA in explaining IL development.

LITERATURE REVIEW

Chomsky (1965) differentiates language competence from language performance. He claims language competence is about language knowledge, while language performance reflects the skills that a learner is able to apply in using that knowledge. Selinker (1972) coins the word *interlanguage* (IL) to refer to the linguistic competence of L2 learners, and since then many researchers have investigated the nature of IL. Tarone (1985) claims that learners showed different IL grammar systems when performing different tasks at the

Published by Australian International Academic Centre PTY.LTD.

http://dx.doi.org/10.7575/aiac.ijalel.v.8n.1p.107

one time. Thus, she questions "the possibility of measuring a learner's grammatical competence" (p.386). Ellis (1987) considers learners' performance as variable competence and concludes: "variability is seen as a feature of the learner's competence, not just of his/her performance" (p. 14). More recently, Tarone (2014) specifies that learners' IL variability is caused by a number of factors, including shifts in social and contextual variables, L1 transfer, and linguistic context. Tarone (1985) and Bayley & Tarone (2012) state IL is systematically and predictably variable across tasks. However, it is not clear whether this variability is meant to refer to either accuracy in performance or acquired knowledge. On the contrary, Pienemann's Steadiness Hypothesis (1998) maintains that the basic nature of the IL system does not change across different communicative tasks provided they are testing the same skill type. Pienemann's Processability Theory (Pienemann 1998) states that the linguistic performance of the L2 learner is determined by their current developmental stage, i.e. their current processing capacity, not by the nature of different tasks (pp. 273-308).

One important element missing in the debate is the meaning of "different tasks". Robinson's (2003, 2005, 2009, 2011) Cognition Hypothesis provides explicit criteria for classifying tasks in terms of (i) task complexity, (ii) task conditions and (iii) task difficulty. Robinson's (2011) Triadic Componential Framework states more cognitively complex tasks may trigger higher levels of accuracy and more complex syntax in performance. Thus, Robinson (2003) states that by definition of "different tasks", effects of task complexity on learner performance (e.g., syntactic complexity, accuracy) could be investigated through other L2 models, such as Processability Theory. No previous studies have tested variability/steadiness of IL systems using tasks of different cognitive complexity.

Robinson's (2003, 2011) criteria for classifying and sequencing pedagogical tasks are both theoretically driven (e.g., Long, 1985, 1998; Merrill, 2006; Reigeluth, 1999; Spector, 2006) and practically researched (e.g., Robinson, Ting & Urwin, 1995; Robinson & Gilabert, 2013). Robinson proposes that it is necessary to distinguish task complexity from task difficulty and task conditions. Robinson (2001a, 2001b) defines task complexity as "the result of the attentional, memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner" (2001b, p.29). Task complexity can help explain within-learner variations. Task conditions include participation and participant factors. For example, a learner's role or status will influence their cooperation and production during interactions. The direction of information flow (e.g., one way vs. two way) and the types of tasks (e.g., one solution vs. many solutions) may affect learners' task performances. Task difficulty refers to the same task potentially leading to different performances among language learners as a result of differences in the attentional, memory and reasoning resources that language learners bring to the task. There are number of variables included in the Triadic Componential Framework. But here in our study, we take up one task complexity variable of "resource-dispersing" +/- planning time.

Robinson (2013) contends that Resource-dispersing variables will not direct learners to any linguistic system, so learners will not acquire any new L2 form-concept mappings. Instead, increasing complexity along resource-dispersing variables will accelerate "automatic access to an already established interlanguage system" (Robinson, 2007, p. 18). In other words, increasing task complexity along resource-dispersing variables makes a task more complex for learners to handle, and learners' accuracy, complexity and fluency decrease at the same time. However, more opportunities are created which enable learners to access more real-time language situations. A learner's ability to use their language knowledge gradually increases as they perform complex tasks with manipulated resource-dispersing variables.

Processability Theory (PT) (Pienemann 1998, Pienemann, Di Biase & Kawaguchi 2005) is a theory of second language processing that formally predicts syntactic and morphological "developmental trajectories" for any given L2. It is therefore assumed to work universally. PT hypothesises that processing procedures and the capacity for the exchange of grammatical information will be acquired in the learner's implicational sequence. PT hierarchy is related to the requirements of the specific procedural skills needed for the target language (any second language). In this way, predictions can be made for language development that can be tested empirically. Table 1 presents the developmental stages of morphology and Table 2 the developmental stages of syntax based on the Lexical Mapping Hypothesis which deals with the association of thematic roles (e.g., agent and patient) and grammatical functions (e.g., subject and object).

RESEARCH DESIGN

Research Question

This study tests the Steadiness Hypothesis on L2 learners' IL systems. The cognition hypothesis is used to test the steadiness of the IL systems. The study investigates if learners' IL systems vary according to tasks of different degrees of cognitive complexity. More specifically, the following questions are asked:

RQ 1. Does the L2 learner's **competence** vary according to the "cognitive complexity" of the tasks? That is, does the L2 developmental stages in syntax as defined by Processability Theory vary according to the variable of +/- planning time?

RQ 2. Does the L2 learner's **performance**, as measured by accuracy, vary according to the cognitive complexity of the tasks according to the variable of +/- planning time?

Hypothesis

Concerning Research Question 1 above, we hypothesise that the learner's L2 competence, as defined by PT, will not vary according to task complexity variable, +/- *planning time*. The claim is based on the Steadiness Hypothesis (Pienemann, 1998), according which learners' IL competence does not change across communicative tasks provided the

Table 1. Processing procedures applied to English L2 morphology (Pienemann, 2005)

Stages	Processing procedures	Information exchange	L2 morphology process	Examples
5	S-procedure	Info exchange within sentence	Subject-Verb agreement	*Lucy likes noodles.* *My mum takes train every day.*
4	VP procedure	Info exchange within VP	VP agreement	*He will go to school.* *She is singing.*
3	NP procedure	Info exchange within NP	NP agreement	*Mum hold two books.* *I have many friends.*
2	Category procedure	No info exchange	Form variation	*Cats under chair* *Driver looking newspaper.*
1	Word/Lemma	Word access no info exchange	Single words, formula	*Fish* *Hello!* *What's your name?*

Table 2. Developmental stages for English syntax based on the association of thematic roles and grammatical function (Pienemann, Di Biase & Kawaguchi, 2005, p. 246)

Stage	Structure	Examples
4. Nondefault mapping	passive, causative, etc.,	*The lamb was eaten by the wolf* *The boss let the workers work from morning to night*
3. Lexically non-default mapping	exceptional verbs	*Lucy puzzled her mother*
2. Default mapping	e.g., agent-event-patient; experiencer-event-theme & canonical word order	*Lucy dancing* *Peter played game*
1. Lemma access	single words; formulas	*Fish* *Hello!* *What's your name?*

tasks test the same skills. As for Research Question 2, it is hypothesised that an L2 learner's performance of tasks (in terms of rule application and grammatical accuracy) will vary with the complexity of the tasks. When the two tasks were manipulated in relation to the +/- planning time variable, the L2 learners tend to produce a greater number of expected structures and had higher VP accuracy rates in the self-paced picture description task (+ planning time) than in the time-defined picture description task (- planning time). This prediction is based on the Cognitive Hypothesis (Robinson, 2011).

Informants

Our informants are 30 college and university Chinese L1 students of English L2 in Australia. These are recruited from three proficiency groups according to their IELTS score, with 10 learners in each group: Low Intermediate (IELTS 4.5-5.0), Intermediate (IELTS 5.5-6.5) and High (IELTS 7.0 or above) proficiency students. All students are overseas students from China. These are code-named as L1-L10, M11-M20 and H21-H30 where L represents "Lower Intermediate", M "Intermediate" and H "High" proficiency students.

Tasks

Two experimental tasks on active/passive alternations were utilised. These tasks were designed by manipulating one of the resource-dispersing variables, that is, the ± *planning time*, which referred to the amount of planning time the tasks involved.

The Self-paced Picture Description Task (+ Planning Time)

This task is considered to be a cognitively less demanding task than the time-defined picture description task (- *planning time*). With the self-paced picture description task, the informants had no time limit and could describe the picture presented on the computer screen and then proceed to the next picture (by pressing the space bar) at their own pace in this task. Microsoft Office PowerPoint 2007 was used to present the stimuli. The informants were instructed to use the brightly coloured item as the subject of each produced sentence. When they were satisfied with their current answer, they could press the space bar to see the next slide. In other words, the informants were given enough time to think about their answers before they started to describe each slide, or they could modify their utterances repeatedly until a satisfactory answer was achieved. Two examples of pictures are given in Figure 1 (a & b) and examples of expected speech productions are provided in (1) (a & b) below. There were two practical trials and 30 eventualities where 15 eventualities were agent-cued (where active sentences are expected as in Figure 1a) and another 15 were patient-cued (where passive sentences are expected as in Figure 1b).

Time-defined Picture Description (- Planning Time)

This task is similar to the self-paced picture description task above. But the difference is that time-defined tasks require the informants to describe each event within a nine-second time limit. Some of the picture and animation events used in this task were taken from Wang's (2010) study, while the others were created by the researchers. DMDX Display software[1] (version 4.3.0.1) was used to present the stimuli for the task to the participants. The time-defined picture description task consisted of two practice trials and thirty eventualities. The 30 eventualities contained 15 agent-cued and 15 patient-cued eventualities. In each event, the cued item was coded in bright colours (e.g. red, blue, yellow and so on), while the remaining items were in black, white or grey. The informants were instructed to use one English sentence to describe each event presented on the screen. Also, they were instructed to use the brightly coloured item as the subject of each produced sentence. We followed Wang (2010) for the timing of this time-defined task. Each trial lasted for 9000 milliseconds (ms). The event for each trial appeared on the computer screen for approximately 5000 ms and the screen turned white during the remaining 4000 ms. Once a trial was completed, the screen turned green for 1000 ms. The green screen signalled to the informants that the computer had stopped recording and they were about to view the next event. Figure 2 (a & b) shows two examples of the time-defined picture description tasks and examples (2) (a & b) are expected speech productions.

Data Analysis

We conducted both competence analyses and performance competence analyses as follows.

Competence measurement is based on Processability Theory's syntactic stages on the Lexical Mapping Hypothesis (see Table 2 above). Productions of active-voice and passive-voice constructions according to the cue are counted to establish each learner's stage. That is, active-voice constructions are expected with agent-cues and passive-voice is expected with patient-cues. In Processability Theory, in order to be qualified to have attained a certain stage, the learner has to produce the structure of that stage more than once involving different lexical choices, such as different nouns and/or different verbs.

Performance analyses in our study is based on active/passive alternation frequencies and the accuracy of the verb phrase (VP). As for accuracy of the VP, different types of errors relating to VP morphology, (3) (a-c) and (4) were counted:

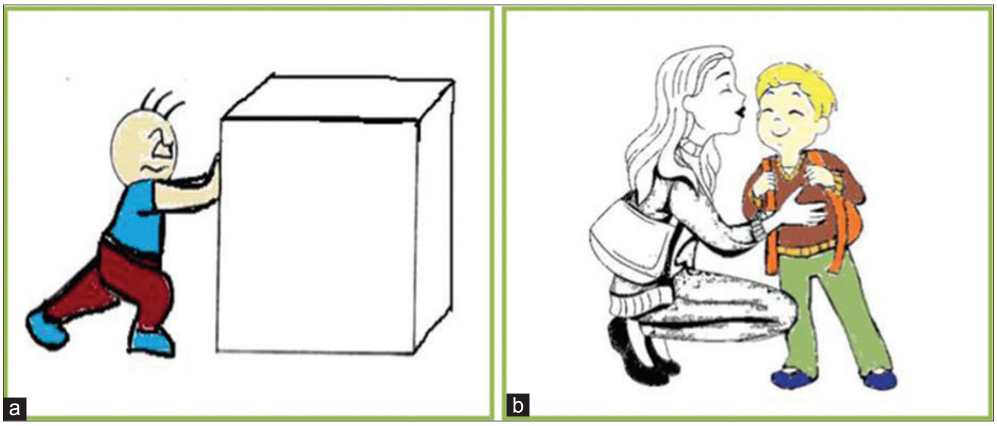


Figure 1. Two examples of the self-paced picture description task. (a) A boy is pushing a cube. (b) The boy is kissed by a woman

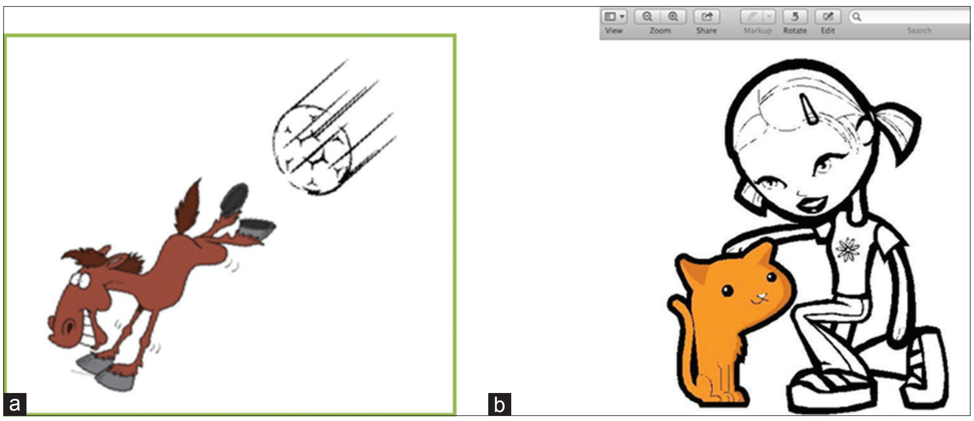


Figure 2. Two examples of the time-defined picture description tasks. (a) The horse kicked/is kicking the ball. (b) The cat is/was patted by the girl

(3) VP agreement errors relating to the auxiliary verb
   a. omission of the auxiliary verb, e.g., *the grey fish _ eaten by white fish*
   b. selection of the wrong auxiliary verb, e.g., *the green fish has eaten by a pink fish* (should be *the green fish is/was eaten)*
   c. the wrong verb form, e.g., *the woman is kiss by a man*

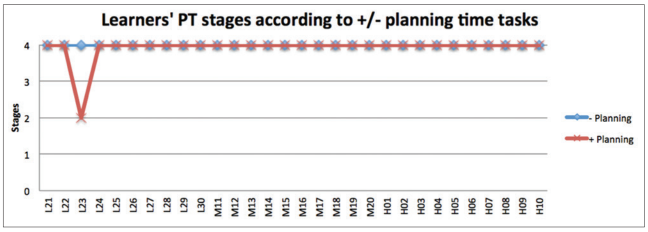
(4) the wrong verb form unrelated to the auxiliary verb, e.g., *the green bottle was hitten by ball*

RESULTS AND DISCUSSION

In order to test if the learner's IL system varies or not, we conducted competence analysis and accuracy (i.e. performance) analysis.

Competence Analysis

Competence analysis tests Research Question 1. We investigated whether the learner's PT developmental stage (See Table 2) varied when one performed tasks of different task complexities, i.e. +/- planning time. Figure 3 presents the 30 learners' PT stages with the tasks of +/- planning time. The results showed that all learners, except L23, were Stage 4 in PT when they performed both +/- planning time. Only L23 showed different stages across tasks: she was Stage 4 when performing the -planning task but she was Stage 2 when performing the +planning time task which is more complex in terms of the cognitive demand imposed on the task, which may be within an error margin.



Performance Analysis

Performance analysis is conducted in order to test Research Question 2. L2 learners' performances are measured by active/passive alternation frequencies and VP accuracy rates.

Active/Passive alternation frequencies

Each task involved 15 agent-cue and 15 patient-cue eventualities. Active voice constructions are expected with agent-cue eventualities and passive voice constructions are expected with patient-cue eventualities. Figures 3, 4 and 5 show the

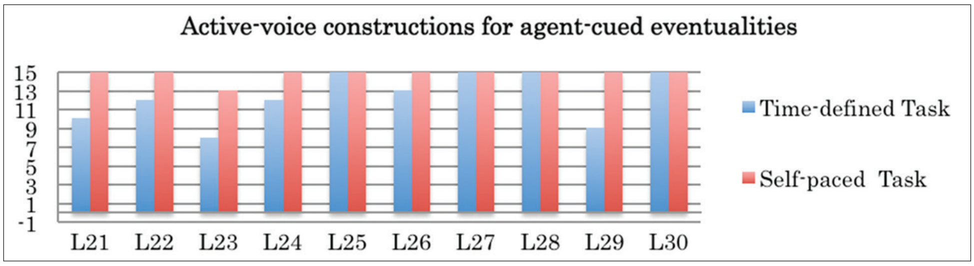


Figure 3. Agent-cue eventualities with Lower-Intermediate Learners

Figure 4. Agent-cue eventualities with Intermediate Learner

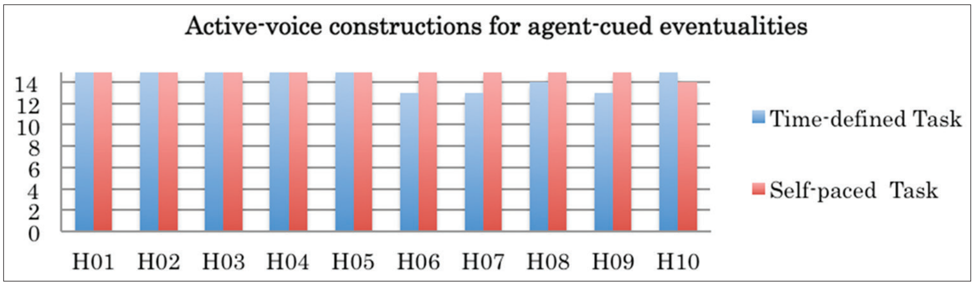


Figure 5. Agent-cue eventualities with High Learner

results of agent-cue eventualities with L (Lower-Intermediate), M (Intermediate) and H (High) learners respectively. Figures 6, 7 and 8 show the results of patient-cue eventualities with L, M and H learners respectively.

Agent-cue eventualities

Almost all learners, regardless of their levels, were able to produce pragmatically appropriate, active voice constructions with the self-paced task (i.e., +planning time) as shown in Figures 3-5. Out of 15 cases of agent-cued eventualities, most learners were able to produce 15 active-voice constructions. The exceptions are L23 (13 cases) and H30 (14 cases). On the other hand, the learners showed some differences with the time-defined task (i.e., -planning time) according to the level of the learner. Some learners of both Lower-Intermediate and Intermediate showed lower performance with the time-defined task compared to their performance with the self-paced task. These learners missed the pragmatic cues due to the time constraints and produced pragmatically inappropriate sentence constructions and/or simply missed describing the eventuality altogether. For example, L23 (in Figure 3) and M14 (in Figure 4) pragmatically appropriate constructions to describe the eventualities only eight and 10 times respectively out of 15. High learners' performance with time-defined tasks tended to be better than Lower-Intermediate and Intermediate learners (see Figure 5). Six out of 10 High learners scored full (i.e., 15 out of 15) producing pragmatically appropriate, active voice constructions. The other four learners scored 13 or 14 respectively out of 15.

Intermediate

Patient-cued eventualities

Passive constructions are expected to describe patient-cued eventualities. The production of passive voice structure (i.e., pragmatically appropriate when the patient is cued) is placed at Stage 4 in PT which is higher than active voice structure (i.e., pragmatically appropriate when agent is cued) which is at Stage 2 in Processability Theory. The results of the patient-cued eventualities turned out to be quite different from the results of the agent-cued eventualities. As can be seen with Figures 6, 7 and 8, Lower-Intermediate learners showed the greatest differences of performance according to the +/- planning time variable (i.e., self-paced and time-defined tasks). The majority (i.e., eight out of ten) of Lower-Intermediate learners performed better with the self-paced task, while two learners (L26 and L28) did better with the time-defined task. These learners commented after the experiment that they had difficulties in keeping up with the cues in the time-defined task and that they sometimes produced active or passive constructions at random. L23 performed especially poorly with both the self-paced and time-defined tasks and she was able to produce only four passive constructions with

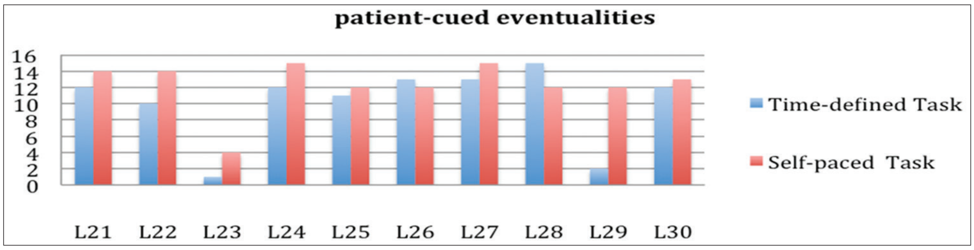


Figure 6. Patient-cued eventualities with Lower Intermediate learners

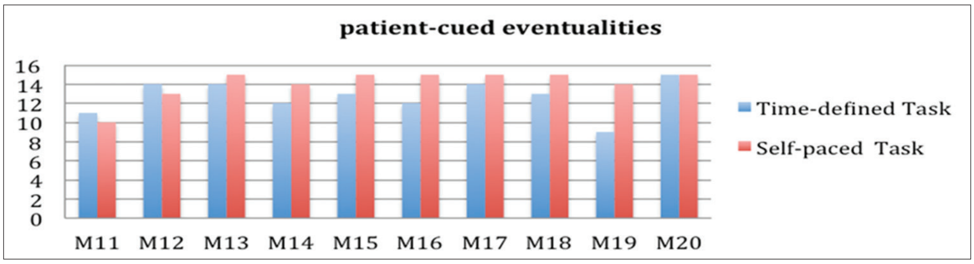


Figure 7. Patient-cued eventualities with Intermediate learners

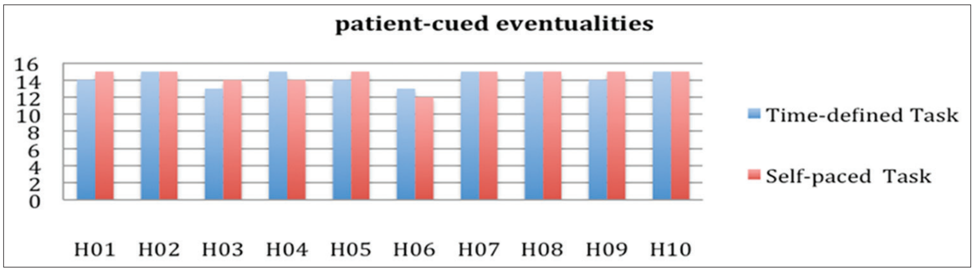


Figure 8. Patient-cued eventualities with High learners

the self-pace task and only one passive construction with time-defined tasks. Intermediate learners performed better than Lower-Intermediate learners (except M19) and were able to produce pragmatically appropriate sentence constructions, that is passive constructions, with the self-paced task. Eight out of 10 Intermediate learners performed better to produce passive constructions with the self-paced task than the time-defined task (see Figure 7). High learners performed much better than both Lower Intermediate and Intermediate learners as all of them, except M11, were able to produce passive constructions 14 or 15 times with the self-paced task. However, even for High learners, their performances with the time-defined task were not as good as the performances with self-paced task (see Figure 8).

Figure 9 shows the summary of frequencies of active voice constructions according to learner proficiency levels with agent-cue eventualities and Figure 10 shows passive voice constructions with patient-eventualities. The differences in active/passive production frequencies between the self-paced task (+planning time) and the time-defined task (-planning) with both patient-cued and agent-cued eventualities became much smaller as the learner's level became higher. This indicates that the linguistic performances become more stable as the proficiency level increases regardless of the tasks of different cognitive complexities. This result is consistent with Segalowitz's (2003) statement: "if one can handle the phonology and syntax of a second language automatically, then more attention can be paid to processing semantic, pragmatic, and sociolinguistic levels of communication".

VP accuracy analysis

VP accuracy rates were analysed based on the criteria listed in Section 3. The accuracy rates were calculated according to the two tasks of the +/- planning time variable.

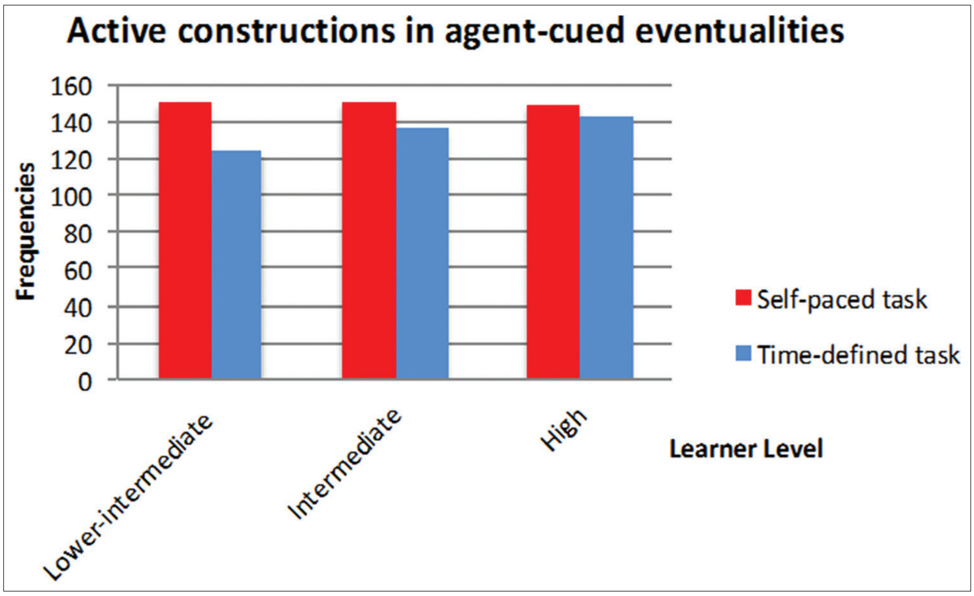


Figure 9. Agent-cued eventualities according to the level of learners

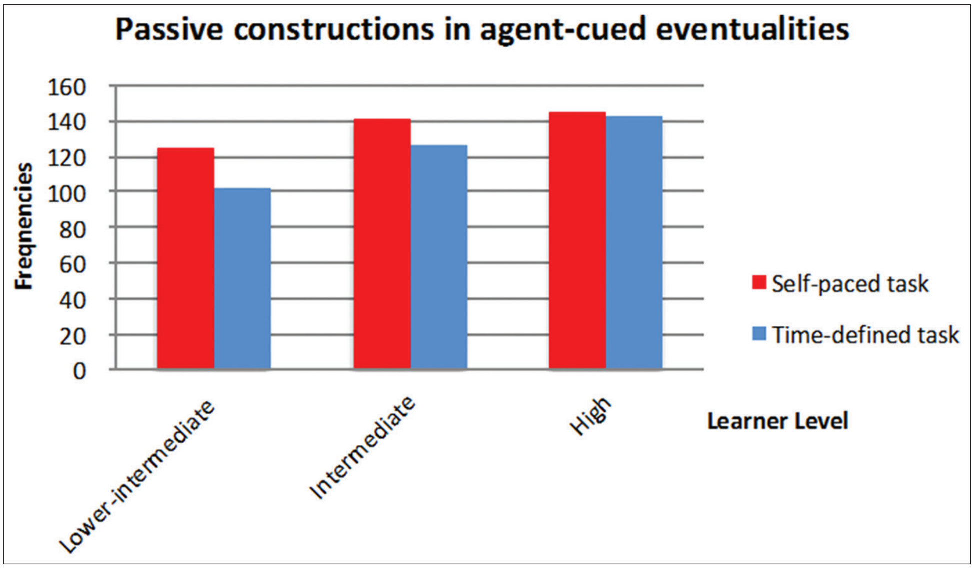


Figure 10. Patient-cued eventualities according to the level of learners

Agent-cue eventualities

Figures 11, 12 and 13 showed VP accuracy rates of the Low-Intermediate, Intermediate and High learners respectively when agent was cued, i.e. active voice constructions are expected (Stage 2 in PT). As can be seen, most learners, regardless of their levels, VP accuracy rates were different with the two tasks of +/- planning time. As for Low-Intermediate learners, the majority (i.e. eight out of 10 learners) performed better with the +planning task than the -planning task. As for Intermediate and High learners, around the half (five and six out of 10 learners respectively) performed better with the +planning task than the -planning task. High learners performed very accurately with both tasks except H06 who achieved only 0.5 with -planning task 1.0 with the +planning task.

Patient-cue eventualities

Figures 14, 15 and 16 represent the accuracy rates of Low Intermediate, Intermediate and High learners respectively when patient was cued, i.e., passive voice constructions are expected (Stage 4 in PT). Here we can see that most learners showed different accuracy rates when they performed two tasks of different cognitive demands, i.e., +/- planning time. VP accuracy rates with Low-intermediate learners were especially lower than that with both Intermediate and High learners.

Table 3 summarises the average VP accuracy rates of agent-cue and patient-cue eventualities respectively with +/- planning time tasks according to the learner levels. As can be seen, it is evident that the accuracy rate improved according to the level of the learners with both active-cue and

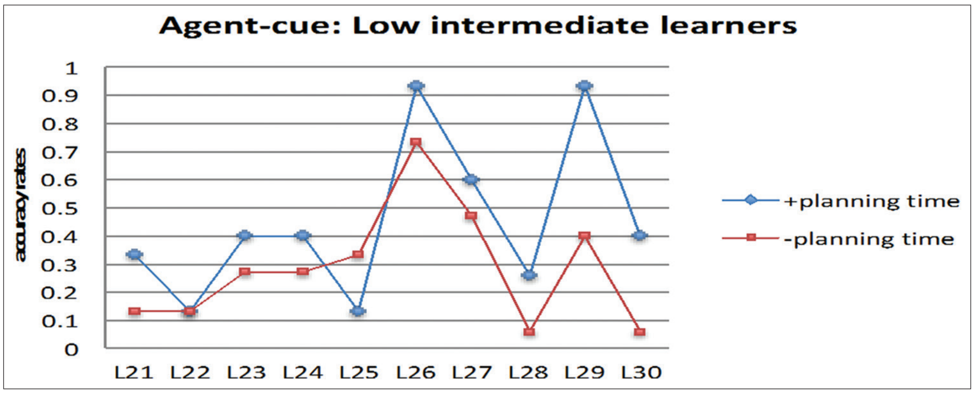


Figure 11. Accuracy analysis of VP constructions: Low intermediate learners

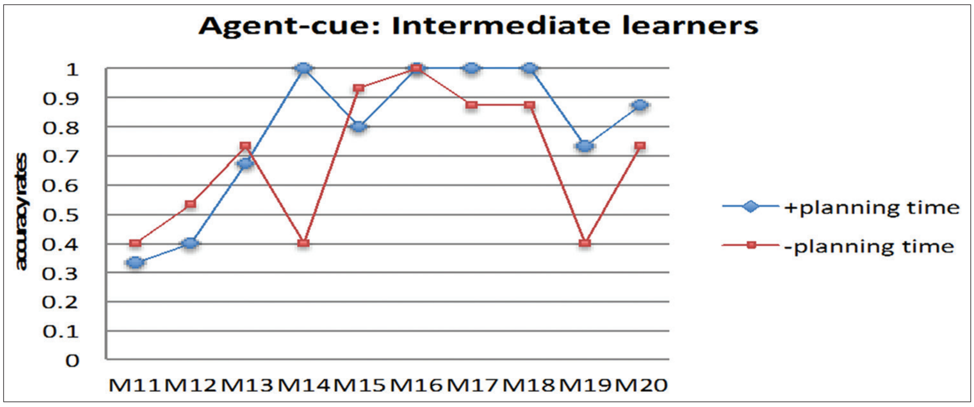


Figure 12. Accuracy analysis of VP constructions: Intermediate learners

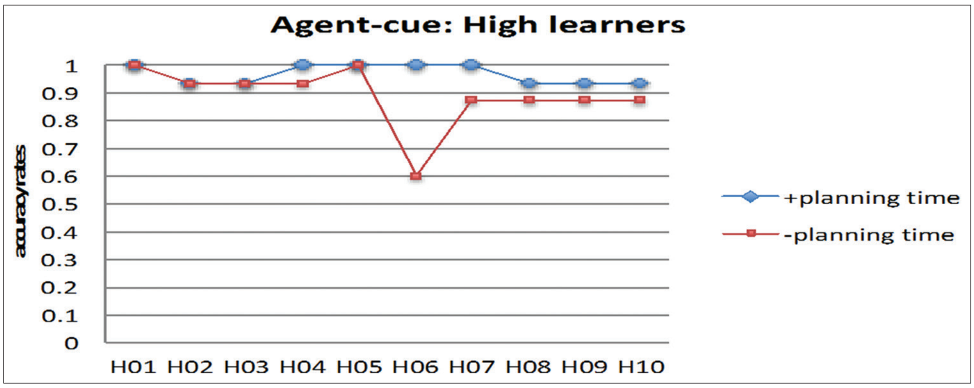


Figure 13. Accuracy analysis of VP constructions: High learners

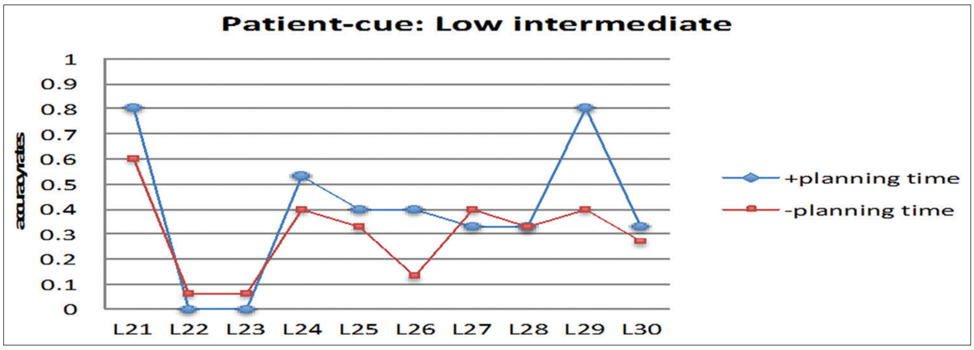


Figure 14. Accuracy analysis of VP constructions: Low-Intermediate learners

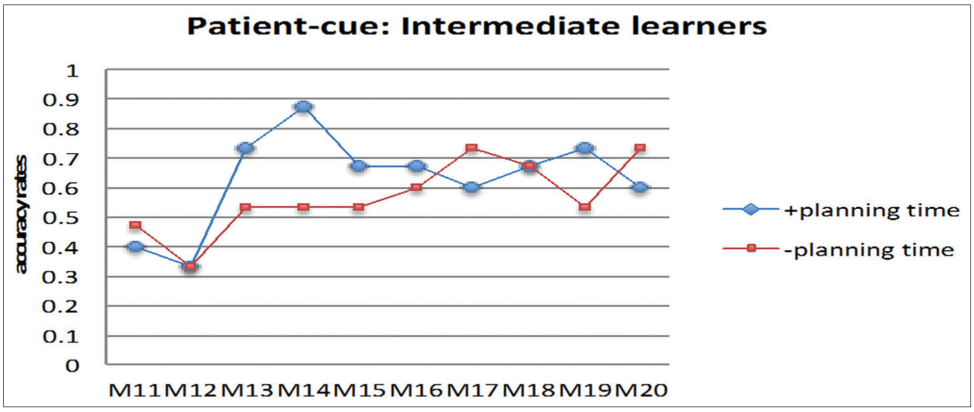


Figure 15. Accuracy analysis of VP constructions: Intermediate learners

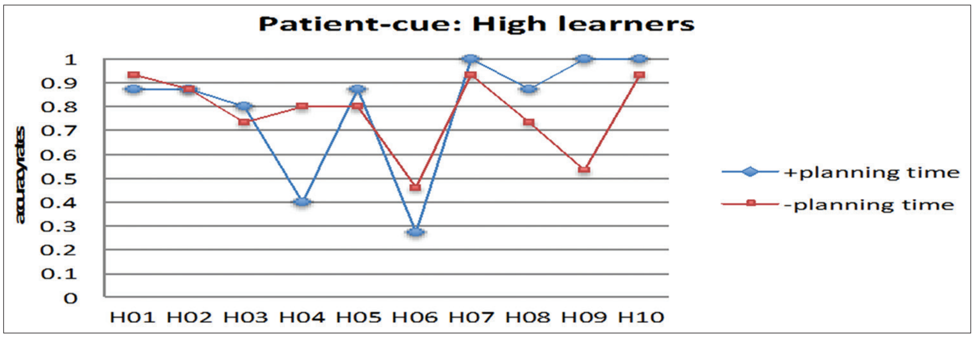


Figure 16. Accuracy analysis of VP constructions: High learners

Table 3. Summary on VP accuracy rates according to the learner levels

Learner levels	Active constructions		Passive constructions	
	Average accuracy rate		Average accuracy rate	
	+ planning time	-planning time	+planning time	-planning time
Lower-intermediate	0.45 (SD=0.29)	0.29 (SD=0.21) *p=0.023	0.39 (SD=0.27)	0.30 (SD=0.17) p=0.124
Intermediate	0.78 (SD=0.25)	0.69 (SD=0.24) p=0.234	0.63 (SD=0.16)	0.57 (SD=0.12) p=0.253
High	0.97 (SD=0.04)	0.89 (SD=0.11) p=0.051	0.80 (SD=0.25)	0.77 (SD=0.17) p=0.634

patient-cue eventualities. For example, Lower-Intermediate, Intermediate and High learners' average accuracy rates with +planning time task for agent-cue eventualities are 0.45, 0.78 and 0.97 respectively. As for patient-cue eventualities, Low-intermediate, Intermediate and High learner's average accuracy rates with +planning task are 0.39, 0.63 and 0.80 respectively. Similar tendencies were found with -planning time task. Dependent (paired) T-tests were carried out to examine if each level of learner showed any difference in performing the two tasks of different cognitive complexities, i.e. +/- planning time. The T-test was carried out according to agent-cue and patient-cue eventualities. Individual scores can be found in Figures 11-16. We found an interesting result. All T-test results found that there was no significant difference between the two tasks of +/- planning time except for Low-Intermediate learners with agent-cue eventualities. Low-Intermediate learner's performances in describing agent-cue eventualities with two tasks were significantly different (*p=.023). This means that Low Intermediate learners performed significantly better when planning time was given. For other learner groups, differences on accuracy rates between the two tasks, +/- planning time, were not statistically significant. Nevertheless, VP accuracy rates with + planning time task were always higher than that with – planning time task across groups. For example, with active constructions, Intermediate learners showed 0.78 accuracy rates with +planning time task while 0.69 with – planning time task. Also, with passive constructions, Intermediate level learner achieved 0.63 accuracy rates with + planning time task but only 0.57 with – planning time task.

CONCLUSION

The present study tested Pienemann's steadiness hypothesis (Pienemann, 1998), which states that the basic nature of the interlanguage system does not change across different communicative tasks, provided they are testing the same skill type. In order to investigate this hypothesis, we asked the following research question: do learners interlanguage systems vary with tasks of different degrees of cognitive complexity? Two tasks were designed by manipulating the task complexity variable, *+/- planning time* that is listed in Robinson's Triadic Componential Framework (Robinson, 2011). Our results suggest that L2 competence, as measured by the developmental stages of Processability Theory, does not appear to vary according to the cognitive complexity of the variables. Therefore, L2 learner's developmental stages remain stable regardless of the tasks with different cognitive complexities. This is compatible with the Steadiness Hypothesis. By contrast, L2 performance as measured by active/passive alternation frequencies and accuracy rates does vary according to cognitive complexity. Tasks of higher cognitive demand showed more varying performances among the L2 learners. Also, lower L2 level learners showed more variable performances and higher L2 level learners were more accurate as well as more stable in their performances. The results of this study thus support Pienemann's Steadiness Hypothesis (1998).

Our results have theoretical and pedagogical implications. As for the theoretical imprecations, our study: a) contributes to Processability Theory by explicitly including task complexity variables in the design of elicitation procedures; and b) adds evidence to Robinson's Cognition Hypothesis from the learners' language developmental perspective. Regarding pedagogical implications, the results of our study offer well-grounded indications for L2 educators, teachers, tasks and syllabus designers in terms of possible effects of task classification criteria and task types on learners' performance. Our study, however, tested only one variable, that is +/- planning time. Therefore, more studies are needed to test other variables affecting the cognitive complexity of the communication task.

In conclusion, this study contributed to the Interlanguage Hypothesis by exploring the nature of a possible IL linguistic system by taking Robinson's task classification criteria into consideration. In particular, IL variability was found in learners' performances, but their IL competence remained steady when the tasks they performed were manipulated using complexity variables.

ACKNOWLEDGEMENTS

An earlier version of this paper was presented at the PacSLRF (The Pacific Second Language Research Forum) in 2016 at the Chuo University, Tokyo. Comments and advice from the floor helped us improve the paper. We are also thankful for the anonymous reviewers for their critical comments and Ms Stacey Sherwood for proofreading the paper.

END NOTES

1. It is stated on the DMDX main webpage: "DMDX is a Win 32-based display system used in psychological laboratories around the world to measure reaction times to visual and auditory stimuli. It was programmed by Jonathan Forster at the University of Arizona".

REFERENCES

Bayley, R., & Tarone, E. (2012). Variationist perspectives. In A. Mackey & S. Gass (Eds.), *Routledge Handbook of second language acquisition* (pp. 41-56). New York: Routledge.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Crystal, D., Fletcher, P., & Garman, P. M. (1976). *The grammatical analysis of language disability*. London: Arnold.

Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition, 9*(1), 1-20.

Long, M. H. (1985). Input and second language acquisition theory. In S. Gass & G. Madde (Eds.), *Input in second language acquisition* (pp. 377-393). California: Newbury House.

Long, M. H. (1998). SLA: Breaking the siege. *University of Hawai'i working papers in ESL, 17*(1), 79-129.

Merrill, M. D. (2006). Hypothesized performance on complex tasks as a function of scaled instructional strategies. In J. Elen & R. E. Clark (Eds.), *Handling complexity in learning environments: Research and theory* (pp. 265-282). Oxford: Elsevier.

Pienemann, M. (1998). *Language processing and second language development: Processability Theory*. Amsterdam: John Benjamins.

Pienemann, M. (2005). *Cross-linguistic aspects of processability theory*. Amsterdam: John Benjamins.

Pienemann, M., Di Biase, B., & Kawaguchi, S. (2005). Extending Processability theory. In M. Pienemann (Ed.), *Cross-linguistic aspects of Processability Theory* (pp. 199-252). Amsterdam: John Benjamins.

Reigeluth, C. M. (1999). What is instructional-design theory and how is it changing? In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory* (Vol. 2, pp. 5-29). Mahwah, NJ: Erlbaum.

Robinson, P. (2001a). Task complexity, cognitive resources, and syllabus design: a triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287-318). Cambridge, UK: Cambridge University Press.

Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*, 27–57.

Robinson, P. (2003). The Cognitive Hypothesis, task design and adult task-based language learning. *Second Language Studies, 21*(2), 45-107.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching, 43*(1), 1-32.

Robinson, P. (2007). Criteria for classifying and sequencing pedagogic tasks. In M. P. Garcia-Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7-26). Clevedon; England: Multilingual Matters.

Robinson, P. (2009). Syllabus design. In M. H. Long & C. Doughty (Eds.), *The handbook of language teaching* (pp. 294-310). Oxford: Blackwell.

Robinson, P. (2011). Second language task complexity, the Cognition Hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second Language Task Complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 3-37). Amsterdam; Philadelphia: John Benjamins Publishing Company.

Robinson, P. (2013). The Routledge encyclopedia of second language acquisition. New York: Routledge.

Robinson, P., & Gilabert, R. (2013). Task-Based Learning: Cognitive Underpinnings. The Encyclopedia of Applied Linguistics.

Robinson, P., Ting, S. C. C., & Urwin, J. J. (1995). Investigating second language task complexity. *RELC journal, 22*(2), 62-79.

Segalowitz, N. (2003). Automaticity and Second Languages. In C. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 382-408). Malden: Blackwell.

Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching, 10*(1-4), 209-232.

Spector, J. M. (2006). From learning to instruction: Adventures and advances in instructional design. In G. Clarebout & J. Elen (Eds.), *Avoiding simplicity, confronting complexity: Advances in studying and designing (computer-based) powerful learning environments* (pp. 15-26). Rotterdam: Sense Publishers.

Tarone, E. E. (1985). Variability in interlanguage use: A study of style-shifting in morphology and syntax. *Language Learning, 35*(3), 373-403.

Tarone, E. E. (2014). Enduring questions from the Interlanguage Hypothesis *Interlanguage: Forty years later* (Vol. 39, pp. 7-26). Amsterdam: John Benjamins

Wang, K. (2010). *The acquisition of English passive constructions by Mandarin speakers: A developmental perspective.* (PhD), University of Western Sydney.