TESOL International Journal

Teaching English to Speakers of Other Languages

Volume 10 Issue 2 2015

ISSN 2094-3938

TESOL INTERNATIONAL JOURNAL

Volume 10 Issue 2, 2015

Published by the TESOL International Journal

http://www.tesol-international-journal.com

Chief Editor: Dr. Xinghua Liu

ISSN. 2094-3938

TESOL International Journal

Contents

Foreword from the editor

Xinghua Liu

I am happy to welcome readers with a brief introduction of the nine papers in this issue. In the first paper, **Dongmei Cheng** and **Jing Liang** compared the effectiveness of explicit and implicit instruction on Chinese EFL learners' production of compliment responses and they found that both types of instruction were equally effective. In the second, **Pino Cutrone** and **Brien Robert Datzman** investigated the effects of a short-term (three-week) study abroad program on a group of Japanese EFL university students' L2 proficiency. They found that the study-abroad students did not outperform their counterparts at home in TOEFL PBT tests. Participants' attitudes and concerns during the program were also discussed. **Yu-ju Hung** presented in the third paper the practice of differentiated instruction with second-graders in an EFL classroom in Taiwan and found that this approach was beneficial for students' language learning with the majority of students perceiving the learning experience positively. In the fourth paper, **Mohammad Tajbakhsh** and **Nasser Gharakhanloo** studied the effects of textual enhanced grammar teaching in comparison to traditional explicit grammar teaching on students' achievement and recall of grammatical structures. It was found that students in the explicit grammar explanation group performed better than those who received textually enhanced materials. The interrelationship between instruction methods and participants' individual differences was also examined. In the fifth paper, **Martin Hawkes** investigated the effects of listening to pre-task models upon priming before students perform meaning-focused tasks in a basic English communication course for first year non-English majors at a Japanese university. The findings demonstrated that pre-task modelling activities positively affected students' strategies and patterns of interaction, and thus suggested that pre-task modelling may be a useful option for teachers using task-based language teaching (TBLT) in their classrooms.

In the sixth paper, **Xinling Zhang, Yan Zhou,** and **Siyu Zhang** studied Chinese EFL learners' writing strategy use in reading-to-write and reading-listening-writing integrated tasks and they found that the majority of participants used discourse synthesis strategy, self-regulatory strategy, and test-wiseness strategy in both tasks during the writing process. **Salah Alfarwan,** in the seventh paper, analyzed narrative patterns and some of the cohesive features found in English narrative compositions produced by Saudi university English majors. Both proficiency-related patterns of organizational features and an L1 transfer effect upon these students' English compositions were found. In the eighth paper, **Attapol Khamkhien** studied the rhetorical and linguistic features of the Discussion Sections of Applied Linguistics research articles which were selected from a Thai research article database. The analysis identified three rhetorical moves of this particular section, namely 'Move 1: Reviewing the present study,' 'Move 2: Analysing and extending from results,' and 'Move 3: Evaluating the study'. In the last paper, **Kyung Sook Cho** and **Stephen Krashen** offered a case observation of how an adult Korean EFL learner developed her English by engaging in self-selected English reading materials over a course of four years and eight months. They argued in support of the positive impact of free voluntary reading on language development.

I would like to take this opportunity to thank my editorial team. They have been so cooperative and devoted. Without their enthusiasm and professionalism, the publication of *TESOL International Journal* would not be possible. Also, I would like to extend many sincere thanks to our external reviewers. Without the work of these generous professionals, a quick and high-quality reviewing process would be impossible. These external reviewers are: Brian Rugen, Joseph P. Vitta, Daguo Li, Hui Li, Dean Jorgensen, Li Zhang, Jiying Yu, Minh Thi Thuy Nguyen, Jean Kirschenmann, Xiaoling Ji, Jinsong (Jason) Fan, and Xinling Zhang.

Xinghua Liu
School of Foreign Languages, Shanghai Jiao Tong University, China
Email: liuxinghua@sjtu.edu.cn

The Effect of Instruction on Chinese EFL Learners' Compliment Responses

Dongmei Cheng*
Texas A&M University-Commerce, USA

Jing Liang
Nankai University, China

Abstract

This study used a pretest-posttest design to compare the effectiveness of explicit and implicit instruction on Chinese EFL learners' (N=41) production of compliment responses (CRs). A video of naturalistic conversations containing CRs by native English speakers was shown to both instruction groups, along with 180 minutes of in-class instruction. The instruction differed in that the explicit group received metapragmatic information on the target feature while the implicit group did not. The tests consisted of naturalistic role-plays and conversations with a native English speaking interlocutor. Participants' CRs were rated using an analytic rubric containing three criteria: semantic strategies, grammar and vocabulary, as well as pronunciation and intonation. The results showed that both types of instruction were equally effective in facilitating participants' production of CRs. Implications for the use of performance-based assessments in interlanguage pragmatic research and suggestions for L2 instruction are provided.

Keywords: compliment responses; speech acts; Chinese; instruction

Introduction

Second language pragmatics instruction has received sustained interest from researchers who have generally reached a consensus on the effectiveness of instruction on L2 speech acts acquisition (Billmyer, 1990a, 1990b; Bouton, 1994; Liddicoat & Crozet, 2001; Vellenga, 2008; Yoshimi, 2001). According to Rose's (2005) overview on the effect of instruction in second language pragmatics, sufficient evidence has indicated that L2 pragmatic instruction, if conducted properly, assists learners' acquisition of various pragmatic features and is more effective than simple exposure to the target language. Due to their close connection to the social preferences and cultural values of the target language, compliment responses (CRs) were among the first speech acts to be included in L2 pragmatic teaching (Holmes & Brown, 1987) and interventional research (Billmyer, 1990a, 1990b). They were regarded as one speech act particularly worth teaching because of the rich evidence shown by research on the dramatic differences in the behaviors of this speech act between English and other languages (e.g., Al Falasi, 2007; Farghal & Al-Khatib, 2001; Golato, 2002, 2005; Lorenzo-Dus, 2001; Nelson, Al-Batal, & Echols, 1996; Tang & Zhang, 2009; Yu, 2003, 2004).

* *Corresponding author. Tel. 903-886-5260; Email: dongmcheng@gmail.com; Address: Department of Literature & Languages, Texas A&M University-Commerce, Commerce, Texas, 75429-3011, USA.*

The effects of two instructional approaches with different degrees of explicitness/implicitness were widely examined in previous studies, which revealed conflicting findings: While a majority of researchers found that explicit instruction, with metapragmatic knowledge provided by the instructor, better facilitated L2 pragmatic acquisition (House, 1996; Jeon & Kaya, 2006; Nguyen, Pham, & Pham, 2012; Rose & Kwai-fun, 2001; Soler, 2005; Takahashi, 2001; Tateyama, 2001), a few studies showed little or no difference between the long term effects of these two approaches (Koike & Pearson, 2005; Martínez-Flor & Fukuya, 2005) or even the opposite result (Kubota, 1995). The present study examines the effectiveness of these two instructional approaches on CRs by adopting two performance-based assessments—naturalistic role-plays[1] (Tran, 2006) and conversations (Baba, 1996)—in order to test the learners' CRs in naturalistic settings.

Compliment Responses in Naturalistic Settings

Compliment responses (CRs) are complicated speech acts because they are closely connected to the cultural values of human societies. In responding to compliments, native speakers from Asian regions such as China, Taiwan, Japan, and Vietnam have been impacted by different cultural values compared to native speakers of English (Chen, 1993; Baba, 1996; Tran, 2006; Yu, 2004). According to Chen (1993), Leech's Agreement Maxim offered good interpretations for the CRs produced by native speakers from the U.S., who predominantly adopted acceptance strategies, while his Modesty Maxim offered fair justifications for the rejection strategies preferred by native speakers of Mandarin Chinese in responding to compliments. However, since English has been gaining a global language status in China, the differences between Chinese and English CRs are diminishing, exemplified by Chinese speakers' adoption of more acceptance strategies and fewer rejections and self-denigrations (Chen & Yang, 2010). Moreover, CRs are also shown to be closely relevant to different contextual variables, such as gender (e.g., Herbert, 1990) and social distances (e.g., Gajaseni, 1995). These variables, together with cultural differences, have added complexity to the study of CRs.

Although intricate speech acts like compliment responses are prevalent in daily conversations, they are difficult to collect. Despite the efforts made by researchers in collecting natural samples of CRs (e.g., Wolfson, 1983), such a data collection task is extremely time-consuming and is often not practical for classroom-based instructional studies. On the other hand, traditional methods in collecting speech act samples (i.e., written or oral discourse completion tasks/DCTs) save the researchers time in obtaining large quantities of data; however, the DCT data only represents participants' knowledge of speech acts, not production. Keeping the balance between research control and authenticity in speech act production, therefore, is a challenge that many researchers are faced with.

Baba (1996) elicited learners' compliment responses via naturalistic conversations. Participants (English as a second language and Japanese as a second language learners) were recruited to carry out conversations with a native speaker of the target language. However, the real research focus (i.e., CRs) was not revealed to the participants. Similarly, in exploring compliment responses made by Vietnamese learners of English, Tran (2006) designed a naturalistic role-play task, which obtained learners' CR productions under the disguise of a series of communicative tasks (e.g., giving directions) without their being aware that CRs were the research focus. These thoughtful research designs ensured the quantity of the target speech acts because the interlocutors were trained to insert compliments in the middle of the conversations, and participants' corresponding responses were thus recorded. Also, since the participants did not know that CRs were the research focus, their responses were spontaneous, which mirrored CR productions in real-life settings.

Cheng (2011) adapted Tran's (2006) role-play task and used it to compare compliment responses produced by Chinese ESL and EFL learners and native English speakers. The results showed L2 speakers', especially EFL speakers, difficulties in utilizing a variety of response strategies other than saying "thank you/thanks" in all complimenting situations, especially on personal traits (e.g., *It's really thoughtful of you to bring a dish.*), whereas native speakers were skillful in using a range of strategies, such as credit-shifting (e.g., *My pleasure/Thanks for having me over.*). EFL learners often have little exposure to authentic language environments, and this can make target speech act instruction difficult, which often requires authentic input. Thus, it is worthwhile to discover an efficient way via classroom instruction to help learners improve their speech act performance.

Two Instructional Approaches with Different Degrees of Explicitness/Implicitness

Following the suggestions by Jeon and Kaya (2006) and Nguyen et al. (2012), the current study treats the distinction between explicit and implicit instruction as a continuum rather than a dichotomy. Previous interventional studies on interlanguage pragmatics mostly compared the effectiveness of two different teaching approaches, with different degrees of explicitness/implicitness. Explicit pragmatic instruction usually refers to different 'focus on forms' techniques (Nguyen et al., 2012), including meta-pragmatic explanations, structural exercises, and direct corrective feedback from the instructor. Implicit pragmatic instruction, on the other hand, has not been well operationalized in previous studies and has often just been described as pure linguistic exposure without the instructor's supply of target language forms and metapragmatic explanations (Jeon & Kaya, 2006). It often has its 'focus on form' (Nguyen et al., 2012) and emphasizes learners' self-discovery of target pragmatic features.

In most interventional studies of speech acts with a comparison group design, the explicit group outperformed the implicit group after the intervention (e.g., House, 1996; Rose & Kwai-fun, 2001; Soler, 2005; Takahashi, 2001; Tateyama, 2001). Soler (2005) discovered that after 30 hours of instruction time, explicit instruction (i.e., focus on forms instruction) was more helpful than implicit instruction (i.e., focus on form instruction) in promoting written production of requests in dialogue forms by Spanish EFL learners. Takahashi (2001) compared the written DCT results on requests by Japanese EFL learners under four instructional approaches with different degrees of input enhancement, each lasting for 6 hours. Results again revealed that an explicit instructional approach, with the provision of metapragmatic instruction, enhanced input to a higher degree when compared to the three implicit approaches, and thus promoted a more positive learning effect. Studies like the above two, by only incorporating written outcome measures, revealed the learners' competency in or perception of the target pragmatic features but failed to document their performance in real communicative contexts.

More studies examined the effectiveness of different instructional approaches by means of oral, or a combination of oral and written measurements (e.g., House, 1996; Liddicoat & Crozet, 2001; Nguyen et al., 2012; Tateyama, 2001; Yoshimi, 2001). These studies, by adopting oral data collection methods, moved one step further in describing learners' pragmatic performance after instruction.

House (1996) examined the acquisition of conversational routines by advanced EFL learners at a German university. After 14 weeks of instruction, results showed that the explicit group outperformed the implicit group in using more types of routine formulae and discourse strategies in conversational tasks. Liddicoat and Crozet (2001) focused on an interactional norm, "Did you have a good weekend", and taught it to a group of beginning level French students in Australia. After a series of consciousness-raising activities[2] containing both explicit and implicit instruction, learners in their study were shown to be capable of maintaining the French-like content in their role-plays in a delayed posttest after one year. Nguyen et al. (2012) investigated the use of criticism realization strategies by high-intermediate college level EFL learners in Vietnam. After 10 weeks of instruction, results also showed that the explicit group performed significantly better than the implicit groups in all measures, including a written DCT, an oral role-play, and oral peer feedback. In a study on teaching the Japanese formulaic expression *sumimasen*, Tateyama (2001) found that beginning level Japanese-as-a-foreign-language learners who received explicit instruction outperformed those who received implicit instruction. Both types of treatment lasted for 50 minutes, and the learners were tested in a multiple-choice test and a structured role-play test rated by native speakers of Japanese. In another study focusing on Japanese pragmatics, Yoshimi (2001) discovered that after 24 hours of explicit instruction, a group of 3rd-year college-level Japanese learners made significant progress in their overall frequency and accuracy in their use of interactional markers in story-telling tasks. The adoption of conversational tasks and role-plays, as illustrated in the above studies, again moved one step further in helping to describe learners' pragmatic performance.

Despite the fact that a majority of interventional studies have demonstrated support for explicit instruction as more effective in enhancing better pragmatic performance, there have been some exceptions. Koike and Pearson's (2005) study on teaching Spanish suggestions to American students (novice-high/intermediate-low proficiency level in Spanish) yielded different results. In this study, the groups receiving explicit instruction and feedback performed better in the multiple-choice test, while those who received implicit instruction and feedback performed better in the dialogue completion task. For both groups, there was no obvious treatment effect after six weeks (or in the delayed posttest). These results indicated that explicit and implicit instruction might contribute to learners' acquisition of different areas of the target pragmatic feature. The authors attributed the main reason for the implicit treatment effect to the use of question recasts, which might have pushed the learners to elaborate and moderate their suggestions and suggestion responses.

House (1996) discovered that both instructional types failed to assist advanced German learners of English in uttering fluent and target-like responses to conversational routines. This finding indicated that explicit instruction was not a panacea for all pragmatic deficiencies. Similar persistent pragmatic problems were evidenced by Yoshimi (2001), whose learners made little progress in managing structural transitions and using interactional discourse markers in story-telling despite the fact that metapragmatic information was provided. Likewise, in Liddicoat and Crozet's (2001) study, after a combination of explicit and implicit consciousness-raising activities, micro-elements of the targeted pragmatic feature, such as those related to language form, were not successfully acquired by French learners.

Another exception is shown in Kubota (1995) on teaching conversational implicature to Japanese EFL learners. Results showed that after 20 minutes of instruction, the implicit instruction group outperformed the explicit instruction group in a multiple-choice and a sentence-composing test, although the differences disappeared by the time of the delayed posttest. Moreover, Tateyama (2001) cautioned the downside of explicit instruction, which might result in the learners' overemphasis and overuse of the instructed features while ignoring other appropriate features in different settings. The teacher-dominated nature of most explicit instruction methods made them less appealing than more inductive instructional methods, in which learners were given opportunities to discover the forms and usages by themselves.

A combination of explicit and implicit instruction was suggested by Martínez-Flor and Fukuya (2005), who found that after 12 hours of instruction, both explicit and implicit groups of intermediate-level Spanish learners of English were able to make great improvement in producing target-like suggestions with no statistically significant difference in the interventional effect.

The only comparison-group interventional study on compliment responses was conducted by Rose and Kwai-fun (2001), who adopted *inductive* (similar to *implicit*) and *deductive* (similar to *explicit*) instruction and compared results from two experimental groups of English learners in Hong Kong. Results from three different written questionnaire measures, including a self-assessment questionnaire, a DCT, and a metalinguistic assessment questionnaire revealed that the deductive group demonstrated a distinctive advantage over the inductive group in producing target-like forms and usages. These findings were not a surprise since the advantage of explicit instruction was to provide metalinguistic knowledge, which was just what was accessed in the questionnaire tasks. The central issues of how learners' use compliment responses in real communicative settings and which instructional type is more effective to help them realize more target-like pragmatic performance were unfortunately still left unanswered.

The results from previous studies indicate that it was perhaps too early to conclude that explicit instruction was the best approach in facilitating target language pragmatic acquisition. Multiple factors should be considered while making comparisons of different instructional approaches, such as the target pragmatic feature, duration of the instruction, learners' proficiency levels, as well as the test selection and possible test effects.

In short, compliment responses, as a complex speech act involving rich social-cultural elements, deserve more emphasis in current pragmatic research, especially in the acquisition of second language pragmatics. Although pragmatic instruction is considered helpful, previous interventional studies have not provided a conclusive answer in terms of whether more explicit instruction can better facilitate second language learners' interlanguage pragmatic development.

This study aims to examine the effectiveness of instruction with two varying degrees of explicitness/implicitness on Chinese EFL learners' compliment responses in naturalistic contexts. With relatively less exposure to target-language features, EFL learners often have more difficulties in acquiring L2 pragmatic knowledge compared to ESL learners (Cheng, 2011; Yu, 2004). These difficulties faced by EFL learners highlight the importance of examining the effectiveness of classroom pragmatic instruction. The present study is designed to answer the following two research questions:

1. Does pragmatic instruction with different degrees of explicitness/implicitness facilitate Chinese EFL learners' compliment response production?
2. Do the two types of instruction impact learners' production of compliment responses differently?

Methodology

Participants

The initial participant pool included 54 sophomores from a large public university in northern China. Since 13 students did not participate in some parts of the study, the final number of participants was 41, which included 19 females and 22 males. They ranged in age from 19 to 22 years, with a mean of 20.02 (SD=.82). All participants were native speakers of Mandarin Chinese, recruited from a public, university-level English listening and speaking course for non-English majors. The two-credit course was offered every semester, and the class met for 90 minutes once per week. A quick survey of the course syllabi and textbooks used by the participants revealed no trace of information on pragmatics in general, nor is the target pragmatic feature, compliment responses addressed in any of the existing course materials. The participants reported a mean of 9.56 years of previous English instruction (SD=1.37) and majored in 17 different subject areas across four academic disciplines, including language and arts, social and behavioral science, business, and science. Their previous English learning experiences were mostly exam-driven, and communicative activities were rarely emphasized in their classrooms. Out of the 41 participants in total, 38 reported that they were preparing for study-abroad exams such as TOEFL, GRE, or IELTS.

Two classes taught by the same instructor were randomly assigned to receive either explicit or implicit treatment, and only students who volunteered to participate in the tests were included in the actual data collection process. The explicit group contained 22 participants (Male=13, Female=9), while the implicit group had 19 participants (Male=9, Female=10). The general oral proficiency of the participating EFL students was assessed using the Test of Spoken English (TSE) Scoring Guide. The participants scored in three different levels, intermediate (TSE=40), higher-intermediate (TSE=50), and high (TSE=60). The explicit group had a mean score of 49.62 (SD=7.74) and the implicit group had a mean score of 51.07 (SD=6.85).

A female native speaker of American English volunteered to be the interlocutor in this study. She had a Bachelor's degree in Public Relations in the U.S. and had one-year English teaching experience at the same university where this study took place. The interlocutor was invited to the EFL participants' classes to meet and talk with the participants during the week before the pretest.

Design

Prior to this study, a small-scale pilot test was conducted with a group of Chinese EFL learners (N=10) from two non-participating classes taught by the same instructor. The pilot test results showed that the naturalistic conversation and role-play tasks were identical in terms of providing the same amount of compliments and eliciting corresponding responses from the participants.

The current study adopted a pretest-treatment-posttest design to examine the effects of explicit and implicit instruction on Chinese EFL learners' compliment responses. The dependent variable was the learners' CRs, which were assessed by the two researchers following an analytic rubric (See Table 1 on p. 17-18). The independent variable was instruction, which had two levels: explicit and implicit instruction. Both groups watched an instructional video on compliments and compliment responses among friends and classmates on a U.S. campus. Explicit instruction involved the supply of metalinguistic information regarding the target pragmatic features revealed in the video (i.e., a total of 8 types of common compliment responses) and direct corrective feedback from the instructor (See handout in Appendix A). Implicit instruction consisted of communicative activities, which placed the students at the center of the classroom: The students were asked to work in groups and to generate and categorize the different types of compliment responses and their usages after watching the video. The instructor only served as a facilitator to the classroom discussion and provided recasts or clarification requests as implicit feedback (See handout in Appendix B). For both groups, the total treatment time lasted for 180 minutes of two class periods over two weeks.

Materials

Background Survey

Each participant completed a background survey in Chinese, which gathered both personal information and information related to their English language learning experiences. Specific questions regarding the students' contact with 'authentic' English language culture (e.g., "Have you learned about how to carry out different language functions in an English-speaking country, such as making a request or apology?) were also included due to their close relevance to the focus of this study.

Treatment Materials

A video created by the first author was used as the main instructional material in this study. The video consisted of seven naturalistic conversations by native English speakers in their daily lives in the U.S. Each conversation happened in a different setting on a university campus, and the speakers demonstrated giving and responding to compliments with their friends and colleagues as part of informal dialogues. To ensure the naturalness of these dialogues, minimum instruction was provided to the speakers, except for the fact that each pair should include one compliment in one of the four most common situations for people to pay compliments in English and Chinese (i.e., appearance/clothing, ability/work, possessions, and personality) (Holmes & Brown, 1987; Chen, 1993). The video lasted for about 20 minutes and covered a range of topics, such as conference presentations, school life, sports playing, dressing up for a job interview, and personal belongings (e.g., a new car). The speakers were all graduate teaching assistants, who spoke clearly in Standard English at an average speed. Each video clip was shot at least twice, and only the versions with the best sound and picture quality were included in the treatment.

Two different sets of handouts were provided to the explicit and implicit group respectively (See Appendix A and B). For the explicit group, a list of all compliment response types illustrated in the video was provided along with examples (different from the ones in the video). The participants were encouraged to match the responses shown in the video to the types listed on their handout, and the most common types of Chinese compliment responses (as cited in Chen, 1993) were listed along with examples in Chinese. Cultural differences and preferences in compliment responses were discussed in class through these examples. The implicit group was asked to generate the types of compliment responses shown in the video on their own through small group work. A transcript of the video was provided to the implicit group for the participants to check and complete their notes. A similar cultural comparison was carried out in the implicit group; however, rather than providing the participants with explicit examples, they were asked to think of their own examples and to compare them with the examples shown in the video. For both groups, an open role-play practice, including four different scenarios, which corresponded with the four most common complimenting situations shown in the video were assigned to each pair as homework (see Appendix C).

Test Materials

Two sets of tasks, including one naturalistic role-play and one conversation were designed for the pretest and the posttest respectively. The role-play task used in the pretest adopted the same version as in Cheng (2011), and the one used in the posttest is included in Appendix D. The conversational task is displayed in Appendix E.

A naturalistic role-play task was designed to elicit spontaneous production in controlled settings. As Tran (2006) has explained, a naturalistic role-play allows the researcher to implement the targeted pragmatic feature into communicative tasks that closely resemble real life situations and to take easy control of the data collection process. The two naturalistic role-play tasks used in this study were adapted versions of the original one designed by Tran (2006). Each role-play was divided into two situations, in which the participant was asked to carry on communicative tasks with the native-English-speaking interlocutor, such as greetings, giving directions, asking for and offering help, and describing a place, a thing, and a procedure. These multiple tasks served as distractions to keep the participant's attention away from the research focus. Two different sets of instructions were given to each role-play participant and interlocutor. No information was provided to the participants indicating that they would receive compliments; however, the interlocutors were made aware of the fact that they needed to pay four compliments (corresponding to the four most common complimenting situations) at different points during the role-play and that they needed to do so naturally by integrating the compliments into their conversations.

A conversation task was also carried out in conjunction with the role-play. Following Baba's (1996) design, the conversation task in this study used an open-ended framework, in which there was minimum content requirement. The only requirement for the student participant was to bring one favorite object to the research site; the interlocutor was asked to start the conversation and to naturally insert four compliments (corresponding to the four common complimenting situations) during the conversation.

The two sequenced naturalistic tasks generated a total of eight compliments with two compliments in each of the four most common complimenting situations. This provided a more accurate depiction of the participants' compliment response behaviors compared to the use of a single task.

Procedure

This study was carried out over a six-week period. In the preparation stage, the research details, including the test scheduling, materials, and teaching approaches were discussed between the two authors over the phone and via email on a weekly basis for one month before the study. The first author came up with the research design and created all the materials. The second author (also the EFL instructor) provided training to the interlocutor and instruction to the participants. In preparation for each test, the interlocutor was first provided with an instructional manual, including the goal of the task, their assigned role, the settings, and detailed steps from initiating to finalizing the task. Then the interlocutor was asked to go over the tasks together with the second author, who role-played an EFL participant's role and answered all questions and concerns from the interlocutor. The interlocutor was also asked to put a checkmark on each item listed on the instructional manual after covering it and to return the manual to the second author after completing each test in the data collection. It was made clear to the interlocutor that each person-to-person conversation should be no longer than 20 minutes. Technical support on how to operate the audio-recording equipment was also provided in the training sessions.

During the first week of the study, the EFL participants were recruited, and their background information was obtained. The pretest started immediately after collecting the participants' information and lasted until the end of the second week. Instruction was provided to both groups over the following two weeks, followed by a posttest.

An informed consent form was provided to the participants at the beginning of this study. After agreeing to participate, each participant completed a background survey. The EFL instructor put an audio-recorder in the classroom while the participants were doing regular in-class role-play activities to familiarize them with the presence of an audio-recorder.

Before the pretest, instruction handouts were provided to both the participants and the interlocutor during the first week. The EFL instructor went through pretest instructions in detail during class time and answered all questions and concerns from the participants. A separate meeting was held with the interlocutor, who was provided with the written information as well as oral explanations regarding the pretest. A pretest sign-up sheet was created based on the detailed daily schedules provided by the interlocutor. Each participant was asked to find one 20-minute time slot which worked best for him or her in the following two weeks and to show up in the designated meeting place with the instructional sheet and the one object they were required to bring. The posttest followed the same preparation procedure as the one in the pretest.

In each test, each participant met with the interlocutor separately for a maximum of 20 minutes at the prescheduled time in the instructor's office. Half of the participants did the role-play first and then the conversation, while the other half did the two tasks in the reverse order. All tasks were audio-recorded with the participants' permission.

Instructional treatment was provided in two class periods for a total of 180 minutes. For both groups, the first 30 minutes of instructional time was used for the EFL instructor to provide background knowledge on compliments and compliment responses. First, the social functions and frequency of usage related to the targeted speech acts in both American and Chinese culture were discussed. Then, a handout on how to give compliments was distributed to all students. A complimenting formula and steps for giving compliments were introduced to both groups of students. The purpose of this introduction was to ensure that every participant had access to the same background knowledge and basic complimenting expressions in preparation for the follow-up class activities.

The rest of the 150 minutes were targeted at instructing students on how to respond to compliments. The self-made video was shown to both groups first. Then, a handout was provided to the explicit and implicit group respectively. The explicit instruction was dominated by the instructor's explanations, with reference to and playback of specific video clips. The implicit instruction was more student-centered. The instructor facilitated the students' self-discovery and class discussion by playing the video multiple times and providing a video-transcript to the class at the end. Both groups were assigned with the same homework, including four role-plays related with giving and responding to compliments in the four most common complimenting situations. The second class period was used as a practicum for students to act out the role-plays in their pairs. The instructor provided feedback to the explicit group by pointing out the inappropriate compliment responses and discussing possible improvements with the pairs. In contrast, either recasts or clarification requests were given to the pairs in the implicit group as indirect feedback. The pairs were asked to reflect on their own performance and on self-improvements without explicit knowledge provided by the instructor.

Analysis

To assess the students' compliment responses, an analytic grading rubric was developed, following an example given by Ishihara (2010). Three criteria were considered to be crucial in realizing a successful compliment response: semantic strategies, grammar and vocabulary, as well as pronunciation and intonation. Table 1 displays this grading rubric.

The two researchers first experimented with this rubric in a norming session, assessing the compliment responses that emerged from a set of 10 naturalistic role-plays and conversations collected from participants who only showed up in the pretest. Concrete examples were drawn from the coding results of this norming session to facilitate the actual data analysis, as illustrated in Table 2.

Paired sample t-tests were used to examine participants' change from pretest to posttest in producing appropriate compliment responses. An ANCOVA test was used to compare the two groups' compliment response behaviors in the posttest. All data were separately coded and scored by the two authors. Inter-rater reliability obtained by Pearson Correlations was .83 for the pretest and .86 for the posttest. The mean rating scores between the two raters were used in the final analyses.

Effect sizes were calculated to measure the magnitude of effects of each instruction type in regards to time difference (i.e., from pretest to posttest). Cohen's d was selected as the measurement, following the standards in Second Language Acquisition for effect sizes recommended by Oswald and Plonsky (2011), "with d = .40 representing a small effect, d = .70 medium, and d = 1.00 a large effect" (p. 99).

Table 1
Analytic Grading Rubric on Compliment Responses

a. Semantic strategies: 4 points

Score	Criteria
4	The student utilizes appropriate and adequate number of semantic strategies in responding to a compliment.
3	The student utilizes appropriate but inadequate number of semantic strategies in responding to a compliment.
2	The student utilizes some inappropriate semantic strategies or lacks important semantic strategies in responding to a compliment.
1	The student utilizes inappropriate semantic strategies in responding to a compliment.
0	The student does not produce any utterance.

b. Grammar and vocabulary: 4 points

Score	Criteria
4	The student uses accurate grammatical structures and word choices in a compliment response.
3	The student has an average use of grammatical structures and word choices with some errors that do not usually cause confusion/misinterpretations.
2	The student has some problematic use of grammatical structures and word choices that often cause confusion/misinterpretation.
1	The student has serious trouble using grammatical structures and makes word choices that often do not make sense.
0	The student does not produce any utterance.

c. Pronunciation and intonation: 4 points

Score	Criteria
4	The student has comprehensible pronunciation and appropriate intonation in responding to a compliment.
3	The student has generally comprehensible pronunciation and moderately appropriate intonation.
2	The student's pronunciation is sometimes incomprehensible and/or contains somewhat inappropriate intonation.
1	The student has largely incomprehensible pronunciation and/or inappropriate intonation.
0	The student does not produce any utterance.

Table 2
Examples and Rationales of Rating Scores on Compliment Responses

a. Semantic strategies: 4 points

Score	Example
4	Interlocutor: I love your MP4 player! It's very cool! Student: Thank you. My father bought it for me. It's a new style. (Rationale: The student appropriately responded using more than two semantic strategies, including an appreciation token, "thank you" and an explanation, "My father bought it for me. It's a new style.")
3	Interlocutor: I heard you got an essay published in the school newspaper! Congratulations! Student: Thanks. (Rationale: Although the student appropriately responded to the compliment using an appreciation token, "thanks", the conversation ended abruptly due to the short response, which distanced the student from the interlocutor.)
2	Interlocutor: I saw your article published in the newspaper! Your writing is great! Student: Maybe just so so ba. (Rationale: The student's compliment response is a direct translation from Chinese, which is a simple downgrade of the compliment (ended by a particle "ba", transferred from Mandarin Chinese). Although downgrading is acceptable in English compliment responses, it is often used together with an appreciation token, like "thank you" or "thanks". Missing this important semantic strategy makes the response somewhat inappropriate and reflects little awareness of the sociocultural norms in the target community.)
1	Interlocutor: I saw you walking in the door, and I saw your glasses. I love them! Student: Why? (Rationale: This is a rather awkward compliment response. It may be considered inappropriate or even rude here because the student is normally not expected to ask a question, which may show a lack of attention or inadequate sociopragmatic comprehension skills.)

b. Grammar and vocabulary: 4 points

Score	Example
4	Interlocutor: I read it in the newspaper last week. And you did such a great job writing it. It was a lot of fun to read. And it made me laugh so hard. Student: Thank you. It's my honor. (Rationale: This compliment response is correctly formed in terms of grammatical structures and word choices.)
3	Interlocutor: I have to say that I love the jacket that you're wearing. Student: It's one of my friends bought from New York City. (Rationale: This compliment response contains an inaccurate grammatical structure, but it does not usually cause confusion/misinterpretations.)
2	Interlocutor: You are a very hard-working student. Student: Em [pause] not so hard.

(Rationale: This compliment response is a sentence fragment without a clear subject reference. It is somewhat difficult to interpret what the student meant by "not so hard".)

1 Interlocutor: You must be a very hard-working student.
Student: Yeah, but it pays.
Interlocutor: It what?
Student: It pays.
Interlocutor: It pays? What do you mean by that?
(Rationale: The student's word choices, "it pays", may not make sense to the listener and created difficulty for understanding.)

Results

Preliminary analysis of the pretest data was carried out to check if the two groups were comparable at the pretests, as well as if the gender and oral proficiency of the learners in the two groups were similar. Gender has been shown to be a variable which can influence learners' production of compliment responses (Cedar, 2006; Farghal & Al-Khatib, 2001; Lorenzo-Dus, 2001); therefore, an independent sample t-test was first conducted to check if male and female participants performed differently in the pretest or posttest. Results showed no significant difference in the performance of compliment responses between the two gender groups (M= 22, F= 19) in either the pretest ($t_{(39)}$ =1.20, p=. 24) or the posttest ($t_{(39)}$ = .63, p= .54). Differences in target-language proficiency levels have also been shown to impact the production of compliment responses (Yu, 1999), so a simple ANOVA was used to check if the three proficiency groups (intermediate n=11; higher-intermediate n=23; and high n=7) performed differently in the pretest. Results showed that the participants' compliment response production was not significantly different among the three proficiency groups ($F_{(2)}$= .17, p= .85). Finally, an independent sample t-test was conducted to examine if there was a pre-existing difference in the compliment response production between the two instruction groups. The two groups did not significantly differ from each other in compliment response performance in the pretest ($t_{(39)}$=1.71, p= .10). In summary, the preliminary analyses indicated that gender, general oral proficiency, and pretest scores were not significant factors in the learners' performance of compliment responses.

The first research question asked whether instruction facilitated Chinese EFL learners' production of compliment responses. Descriptive statistics showed that the explicit group generated more appropriate compliment responses at the posttest (M=9.93, SD=1.31) than the pretest (M=7.98, SD=1.29). A similar pattern was also shown for the implicit group, with better performance of compliment responses at the posttest (M=10.48, SD=1.09) than the pretest (M=8.80; SD=1.76). This positive trend is illustrated in Figure 1.

To answer the first research question, a paired-sample t-test was conducted for each instruction group separately. Findings showed that there was a significant increase from the pretest to the posttest scores for the explicit group ($t_{(21)}$ =4.75, p=.00) and for the implicit group ($t_{(18)}$ =3.60, p=.00). Cohen's d was computed based on the descriptive statistics, and a large effect size was shown for both the explicit (d= 1.50) and the implicit group (d= 1.15). Therefore, both types of instruction were considered to have a positive impact on the participants' performance of compliment responses.

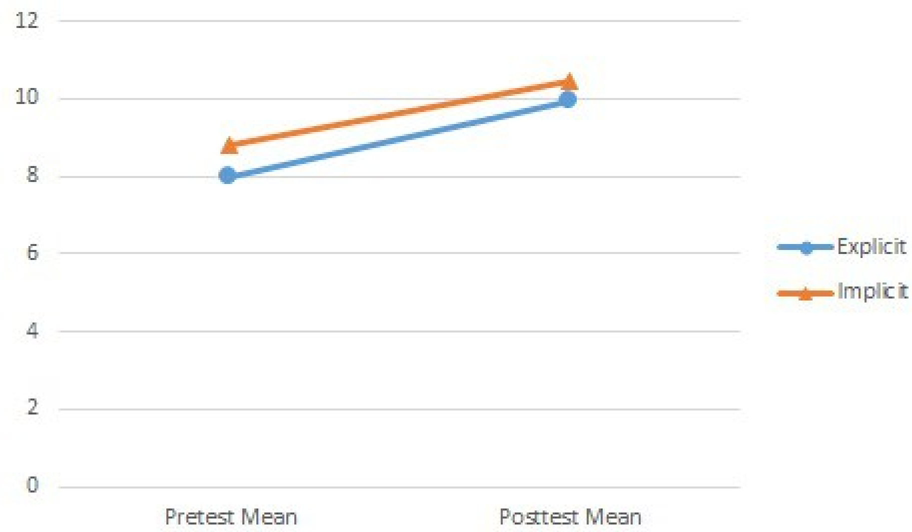

Figure 1. Pretest and Posttest Mean Scores for the Explicit and Implicit Groups.

The second research question asked whether the two instruction types had a different impact on the participants' compliment response behaviors. As previously illustrated in Figure 1, an increased score for compliment responses was observed for both the explicit group and the implicit group. In order to compare the effects of the two instruction types, an ANCOVA test was used with pretest scores included as a covariate. The results showed that there was no significant difference on the post-test scores between the explicit and the implicit group ($F_{(1)}$ =2.06, p=.16).

Discussion

This study explored the effectiveness of instruction and compared the impact of two types of instruction on Chinese EFL learners' production of target-like compliment responses. The results showed that learners had significant improvement in their compliment response performance after instruction and that there was no significant difference between the effectiveness of explicit and implicit instruction. The fact that the explicit group significantly improved after the instruction came as no surprise, being consistent with previous findings (House, 1996; Nguyen et al., 2012; Rose & Kwai-fun, 2001; Soler, 2005; Takahashi, 2001; Tateyama, 2001). The similar positive effect of implicit instruction was in line with the findings by Martínez-Flor and Fukuya (2005), whose study also showed the usefulness of implicit instruction in facilitating L2 learners' pragmatic performance. The different findings on the effectiveness of implicit instruction between the current study and many previous studies may be attributed to many factors, such as the different operationalization of implicit instruction, the different target pragmatic features, as well as the differences in students' level of proficiency of the target language. In the present study, the implicit group had more exposure to the target language in a textual form than the explicit group by having the entire video transcript. This might, to a large extent, have compensated for the lack of metapragmatic information provided by the instructor. In addition, given sufficient consciousness-raising activities provided by both types of instruction, the pragmalinguistic forms of compliment responses, compared to those of many other speech acts (e.g., requests, criticisms, and implicatures), might be easier to acquire for the participants in this study, with all of them having an intermediate-advanced oral proficiency based on their TSE scores.

In contrast to the findings in the majority of previous studies, this study showed no significant difference between the effectiveness of the two instruction types, which resembled the results from Martínez-Flor and

Fukuya (2005), despite the fact that a different learner group and pragmatic feature were focused on in the current study. This result indicates that both explicit and implicit instruction positively contribute to learners' pragmatic production by raising their awareness in producing target-like compliment responses. Compared with the average effect sizes reported in Jeon and Kaya's (2006) meta-analysis, the explicit instruction in the current study generated a smaller effect size (d=1.50) in comparison to the average effect size for explicit pragmatic instruction (d=1.91), while the implicit instruction had only a slightly larger effect size (d=1.15) compared to the average (d=1.01).

The learners' awareness was illustrated in the posttest dialogues of several participants from both groups, who indicated that they were not only aware of the fact that they were given a compliment, but also understood the differences between the typical American ways and the traditional Chinese ways to respond to a compliment. One participant from the explicit group, for example, said to the interlocutor after receiving multiple compliments in the conversation:

> I think foreigners always like to praise people. You're giving me too much praise and sometimes I feel that I almost cannot manage them. We Chinese do not like to praise others too often, because others may think you want something from them if you always say good things.

This excerpt shows that the participant recognized that the interlocutor had paid them compliments, although such awareness was not shown in the pretest data. It is also noteworthy to highlight that the participant was comparing the two cultures on the frequency of giving compliments and the social thoughts as well as expressing their own feelings towards receiving compliments. Such detailed reaction and analysis might be influenced by the activities the students did during the in-class instruction, when they were encouraged to think of typical compliment responses in their home culture and to compare them with the most commonly occurring American compliment responses in the instructional video.

Surprisingly, even without explicit knowledge being provided, participants in the implicit group also demonstrated the same awareness of the compliments they had received in the conversation and the role-play. In the posttest, one participant from the implicit group commented after receiving a compliment from the interlocutor that "*you always like the things that I showed to you*" and expressed their happiness that they both had similar preferences. Again, this type of response was not shown in the pretest data.

In terms of the typical response types produced by the participants, both groups were able to provide more detailed explanations after receiving compliments, just like what was shown by the native English speakers in naturalistic settings (Cheng, 2011). Frequent use of explanations as a compliment response strategy was a more advanced step for the learners to approach target-like performance because providing relevant explanations regarding the complimenting objects or subjects was more cognitively demanding for L2 learners, compared to a simple reaction using formulaic appreciation token forms (i.e., *Thank you*.). The following are two excerpts from the naturalistic role-play task in the pretest and posttest from the same participant, who demonstrated a typical change in the use of explanations.

> 1) Excerpt from the pretest role-play task:
> Interlocutor: Wow, it is such a cool bike! It's so pretty. I like the color.
> Participant: Thank you.
>
> 2) Excerpt from the post-test role-play task:
> Interlocutor: Yours (referring to the MP4 player) looks really cool. It looks like it is made very well, has good technology and I really like the color of it.
> Participant: Thanks. I got this from my father as a birthday gift. I also really love the white color. Not so many are like this in the market.

In both excerpts, compliments were paid to the participant's personal belongings. In the pretest, a typical response that most participants adopted was a simple *thank you*, just like the participant in the first excerpt. Although this response was not wrong or impolite, the conversation usually came to a stop whenever more information was not provided after saying *thank you*. This discontinuation of the talk could potentially weaken the communication forces between the two speakers. Since the conversation and role-play tasks in this study closely resembled real-life communication between friends, a lack of words from one speaker could lead to silence and frustration from the other speaker. The analysis of the interlocutor's conversation scripts showed that they were able to come up with more information and be more engaged in the tasks when the participants produced more explanations in response to their compliments. This improvement in compliment response production was shown in the second excerpt from the posttest, which demonstrated the participant's ability to keep the conversation going by providing explanations after showing appreciation for the compliment.

This study has implications for both interlanguage pragmatic research and classroom pragmatic instruction. The use of two innovative performance-based assessments, naturalistic conversation (Baba, 1996) and naturalistic role-play (Tran, 2006) successfully captured the near-authentic compliment response productions of the participants. Since the content of the conversations and the role-plays was also common in the participants' daily lives, they could easily make connections to real life settings and appreciate the relevance of the tasks to their own goals of producing increasingly target-like English. The fact that a majority of the participants in this study were preparing to study abroad in an English-speaking country after graduation provided more evidence on the potentially motivating influence of adopting these naturalistic tasks in interlanguage pragmatic studies with similar learner groups.

Besides its contribution to research methodology, this study has made useful suggestions on classroom pragmatic instruction, especially for EFL contexts, where class size is usually large and the students have typically gone through years of English training yet have rarely had access to an 'authentic' language environment. The fact that a significant positive effect was observed for both the explicit and implicit group after a relatively short period of instruction indicates that the instruction provided in this study was effective in terms of heightening the learners' awareness of the target pragmatic feature and offering them choices for producing it more appropriately.

The use of authentic video as a major source of input was beneficial in two ways: First, it covered daily conversation topics that were common in the participants' own lives. Secondly, it showed snapshots of real life situations and casual conversations between friends in the U.S. These snapshots were distinct from the movies or TV shows the students often watched. There is usually a difference between the speech styles presented in media and the ways average people talk in real life. Thus, the self-made video in this study was particularly valuable in terms of providing the participants with a short-cut to see and hear authentic complimenting and responding behaviors demonstrated by native English speakers without stepping outside their classrooms.

Furthermore, the activity of comparing the learners' home cultural values and preferences to those in the target culture was effective in raising their awareness of cultural differences, understanding why people in the target culture prefer certain types of compliment responses and learning to accept and appreciate the different responses, which might not be present in their home culture. The fact that the implicit group also improved significantly in the posttest showed that the learners had the ability to identify the most appropriate ways of responding to compliments and realizing the cultural differences on their own, from the same instructional material. This result is particularly exciting because it gives instructors the option to form a student-centered classroom and to give the students more control over their own learning. Either instruction type or a combination of both might be implemented at any time while introducing a new pragmatic feature, especially on certain speech acts which have similar characteristics to the one examined in this study.

Limitations and Further Research

The use of naturalistic tasks in this study helped to generate more authentic responses from participants; however, this authenticity necessarily results in a lack of control. Sometimes the interlocutor failed to provide a compliment or provided more than one compliment in a particular situation. This might also be due to the heavy work load and tight research schedule of the interlocutor in this study, who had to conduct hours of tasks with all participants over two weekends for both the pretest and the posttest. In future studies adopting similar assessment tools, more careful and detailed instruction as well as sufficient time needs to be provided to the interlocutor.

The lack of research control is particularly reflected in the naturalistic role-play tasks. Since no exemplary forms of compliments were provided for the interlocutor in the instructional manual, they could produce different forms of compliments for different participants, which could potentially lead to different types of responses.

In addition, the sample size in this study is relatively small (N=41). All participants were native Chinese speakers in a specific EFL context, and only one pragmatic feature was examined. These facts limit the generalizability of the results to larger populations, to other L2 learners in different instructional settings, and for the use of different pragmatic features. Also, a lack of control group in a quasi-experimental study like this is another drawback, because without comparison to a control group that received no pragmatic instruction, it is impossible for us to know "the absolute effectiveness of the instruction" (Jeon & Kaya, 2006, p. 183). Therefore, future research should consider using a larger sample and a control group as well as examining different pragmatic features.

Conclusion

This study found that both explicit and implicit pragmatic instruction successfully facilitated Chinese EFL learners' production of appropriate compliment responses and that these two instruction types did not differ significantly in terms of impact. These results suggest the potential benefits of combining explicit and implicit instruction in teaching second language pragmatics (Martínez-Flor & Fukuya, 2005). The successful implementation of a self-made video as an instructional material provided insights for lesson plans in teaching speech acts and set up a model for making pragmatic instruction more lively and authentic. This study also showed the benefits of introducing cultural comparisons in the process of pragmatic instruction. Learners understood the authentic language input more thoroughly by examining contrasting examples from their first language culture. Finally, examining learners' production of pragmatic features in real communicative contexts via naturalistic conversations and/or role-plays is recommended for other studies exploring the use of speech acts other than compliment responses (see another example in Nguyen, 2014). This continuous pursuit in task design will surely benefit pragmatic research methodology in general and help in depicting a more accurate and complete picture for learners' development in acquiring other speech acts.

Endnotes

1. Tran (2006) used the term "Naturalized role-play (NRP)".
2. Liddicoat and Crozet (2001) did not specify the duration of their instruction/consciousness-raising activities.

References

Al Falasi, H. (2007). Just say "thank you": A study of compliment responses. *The Linguistics Journal, 2*, 28-42.

Baba, J. (1996). *A study of interlanguage pragmatics: Compliment responses by learners of Japanese and English as a second language*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 9719290)

Billmyer, K. (1990a). "I really like your lifestyle": ESL learners learning how to compliment. *Working Papers in Educational Linguistics, 6* (2), 31-48.

Billmyer, K. (1990b). *The effect of formal instruction on the development of sociolinguistic competence: The performance of compliments.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 9026523)

Bouton, L. (1994). Conversational implicature in the second language: Learned slowly when not deliberately taught. *Journal of Pragmatics, 22*, 157-167.

Cedar, P. (2006). Thai and American responses to compliments in English. *The Linguistics Journal, 2*, 6-28.

Chen, R. (1993). Responding to compliments: A contrastive study of politeness strategies between American English and Chinese speakers. *Journal of Pragmatics, 20*, 49-75.

Chen, R., & Yang, D. (2010). Responding to compliments in Chinese: Has it changed? *Journal of Pragmatics, 42,* 1951-1963.

Cheng, D. (2011). New insights on compliment responses: A comparison between native English speakers and Chinese L2 speakers. *Journal of Pragmatics, 43,* 2204-2214.

Farghal, M., & Al-Khatib, M. (2001). Jordanian college students' responses to compliments: A pilot study. *Journal of Pragmatics, 33,* 1485-1502.

Gajaseni, C. (1995). A contrastive study of compliment responses in American English and Thai including the effect of gender and social status (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 9512365)

Golato, A. (2002). German compliment responses. *Journal of Pragmatics, 34,* 547-571.

Golato, A. (2005). *Compliments and compliment responses: Grammatical structure and sequential organization.* Philadelphia, PA: John Benjamins.

Herbert, R. (1990). Sex-based differences in compliment behavior. *Language in Society, 19*, 201-224

Holmes, J., & Brown, D. (1987). Teachers and students learning about compliments. *TESOL Quarterly, 21,* 523-546.

House, J. (1996). Developing pragmatic fluency in English as a foreign language: Routines and metapragmatic awareness. *Studies in Second Language Acquisition, 18,* 225-252.

Ishihara, N. (2010). Assessing learners' pragmatic ability in the classroom. In D. Tatsuki & N. Houck (Eds.), *Pragmatics: Teaching speech acts* (pp. 209-227). Alexandria, VA: Teachers of English to Speakers of Other Languages.

Jeon, E. H., & Kaya, T. (2006). Effects of L2 instruction on interlanguage pragmatic development. In J. M. Norris. & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 165-211). Amsterdam, the Netherlands: John Benjamins.

Koike, D., & Pearson, L. (2005). The effect of instruction and feedback in the development of pragmatic competence. *System, 33,* 481-501.

Kubota, M. (1995). Teachability of conversational implicature to Japanese EFL learners. *IRLT Bulletin, 9,* 35-67.

Liddicoat, A., & Crozet, C. (2001). Acquiring French interactional norms through instruction. In K. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 125-144). Cambridge, England: Cambridge University Press.

Lorenzo-Dus, N. (2001). Compliment responses among British and Spanish university students: A contrastive study. *Journal of Pragmatics, 33*, 107-127.

Martínez-Flor, A., & Fukuya, Y. (2005). The effects of instruction on learners' production of appropriate and accurate suggestions. *System, 33,* 463-480.

Nelson, G., Al-Batal, M., & Echols, E. (1996). Arabic and English compliment responses: Potential for pragmatic failure. *Applied Linguistics, 17* (4), 411-432.

Nguyen, T.T.M., Pham, T.H., & Pham, M.T. (2012). The relative effects of explicit and implicit form-focused instruction on the development of L2 pragmatic competence. *Journal of Pragmatics, 44,* 416-434.

Nguyen, M. T. T. (2014). Using conversation tasks and retrospective methodology to investigate L2 pragmatics development: the case of EFL criticisms and responses to criticisms. *The Language Learning Journal*. Advance online publication. doi: 10.1080/09571736.2014.908404

Oswald, F., & Plonsky, L. (2010). Meta-analysis in second language research: choices and challenges. *Annual review of applied linguistics, 30,* 85-110.

Rose, K. (2005). On the effects of instruction in second language pragmatics. *System, 33,* 385-399.

Rose, K., & Kwai-fun, C. (2001). Inductive and deductive teaching of compliments and compliment responses. In K. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 145-170). Cambridge, England: Cambridge University Press.

Soler, E. A. (2005). Does instruction work for learning pragmatics in the EFL context? *System, 33*, 417-435.

Takahashi, S. (2001). The role of input enhancement in developing pragmatic competence. In K. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 171-199). Cambridge, England: Cambridge University Press.

Tateyama, Y. (2001). Explicit and implicit teaching of pragmatics routines: Japanese sumimasen. In K. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 200-222). Cambridge, England: Cambridge University Press.

Tang, C., & Zhang, G. Q. (2009). A contrastive study of compliment responses among Australian English and Mandarin Chinese speakers. *Journal of Pragmatics, 41,* 325-345.

Tran, G. Q. (2006). The nature and conditions of pragmatic and discourse transfer investigated through naturalized role-play. Muenchen, Germany: Lincom Europa.

Vellenga, H. (2008). *Instructional effectiveness and interlanguage pragmatics*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3318496)

Wolfson, N. (1983). An empirically based analysis of complimenting in American English. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition.* (pp. 82-95). London, England: Newbury House.

Yoshimi, D. (2001). Explicit instruction and JFL learners' use of interactional discourse markers. In K. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 223-244). Cambridge, England: Cambridge University Press.

Yu, M. (1999). *Cross-cultural and interlanguage pragmatics: Developing communicative competence in a second language.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 9933159)

Yu, M. (2003). On the universality of face: evidence from Chinese compliment response behavior. *Journal of Pragmatics, 35*, 1679-1710.

Yu, M. (2004). Interlinguistic variation and similarity in second language speech behavior. *The Modern Language Journal, 88,* 102-119.

Appendix A
Handout Provided for the Explicit Instructional Group

There are multiple ways to respond to compliments in English. Some of them are more acceptable than others in English-speaking countries:

1. Appreciation token: The complimentee recognizes the compliment and shows appreciation for it. Appreciation token usually does not fit into the meaning of the compliment.

    A: What a lovely dress!

    B: Oh. Thank you. This is one of my favorite.

2. Explanation/Comment History: The complimentee impersonalizes the compliment assertion by giving further information, which may frequently be irrelevant about the object of the compliment.

    A: I like your tie. It suits you well.

    B: Thanks. Mom bought it for me. She likes to buy me nice ties now and again.

3. Compliment upgrade: The complimentee agrees with and increases the compliment assertion.

    A: Nice car!

    B: Thanks. Brand new.

4. Return: The complimentee reciprocates the act of the complimenting by paying back the compliment to the complimenter.

    A: You're looking good.

    B: Thanks. So are you.

    Or

    A: You're looking good today.

    B: Thank you very much. Not too bad yourself.

5. Disagreement: The complimentee directly disagrees with the praise force/compliment assertion. They assert that the praise within the compliment is overdone or undue.

    A: You're such a talented writer. I've learned so much from you.

    B: No, I'm not. In fact I got a lot of help from my advisor.

6. Agreement: The complimentee agrees with the compliment assertion by providing a response which fit meaningfully to the compliment.

    A: Hey you look really nice today.

    B: Yeah I'm happy to say that that's correct. Heh heh heh.

An agreement can be scaled down to minimize the force of the compliment.

    A: I like your car. It's very good.

    B: Oh, Yeah. Thanks. It's not bad.

7. Compliment downgrade: The complimentee qualifies the praise force/compliment assertion, or downplays the object of the compliment.

    A: It's a really nice car.

    B: Oh no. It looks like that but actually it has a lot of problems.

    Or

    A: It's a nice car. I really like it.

    B: Oh well. It's just a normal and not very reliable car.

8. Expressing Gladness: The complimentee does not address the compliment assertion itself, which makes the response a type of avoidance, but expresses their gladness that the complimenter likes the object of the compliment.

    A: By the way, I read your article that you published last week. It was very good.

    B: Oh, that's good. Thank you.

    Or

    A: I read your article you published last week. It was very good.

    B: Well, it's great you say that.

To Compare: Chinese Compliment Responses

- Rejecting (96%)
  1. Disagreeing and denigrating 51%

     A: 你刚才的演讲做得真好! B: 哪有?但愿我没有把大家都讲睡着了.

  2. Expressing embarrassment 26%

     A: 你的歌唱得真不错！ B： 夸得我都不好意思了．．．

  3. Explaining 19%

     A： 这件毛衣很好看呀！ B： 哦，都穿了好几年了.

- Thanking and denigrating
  1. Thanking and denigrating 3%

     A： 你刚才的表演真是太精彩了！ B： 谢谢啦。过奖过奖。

- Accepting
  1. Thanking only 1%

     A: 这道菜做得真好吃！B：谢谢。

Appendix B
Handout Provided for the Implicit Instructional Group

I. While watching the video, please write down the complimenting topic(s) in each scenario:

II. How did the speaker in the video respond to the compliment? Write down the compliment responses in each scenario:

III. Take a look at the above compliment responses (check with the video transcript and generalize how native English speakers of American English respond to compliments in different situations)
1) Appearance/clothing:
2) Possession:
3) Ability/work:
4) Personality:

IV. Compare this to what you would say in Chinese: Make your own Chinese role-plays under the above four complimenting categories. Think about what you would actually say in your daily life and make your-role-play as authentic as possible.

Appendix C
Open Role-Play Practice

With a partner, practice paying and responding to compliments in the following four settings:

I. Your friend has just won first place in the English speaking contest. You two meet and start talking. Your talk should include at least the following three things:

1. Greetings; asking about school/life
2. Complimenting & responses to compliment (on winning the 1st place in the contest)
3. Shifting to another topic (anything that you like to talk about; make sure that you have a smooth transition from the compliment to the new topic; e.g., going to lunch/dinner; inviting to a weekend party; inviting to go to the English corner, etc.)

II. Your friend dressed very formally/pretty today. You two meet and start talking. Your talk should include at least the following three things:

1. Greetings; asking about school/life
2. Complimenting & responses to compliment (on the formal/beautiful clothes or any other special feature of your friend's appearance/clothing)
3. Shifting to another topic (anything that you like to talk about; make sure that you have a smooth transition from the compliment to the new topic)

III. Your friend has just bought a new bag/cellphone/MP4 player (or any object that you like to use). You two meet and start talking. Your talk should include at least the following three things:

1. Greetings, asking about school/life
2. Complimenting & responses to compliment (on the new object bought by your friend)
3. Shifting to another topic (anything that you like to talk about; make sure that you have a smooth transition from the compliment to the new topic)

IV. Your friend has just come back from Wenchuan as a volunteer to help with reconstruction. You two meet and start talking. Your talk should include at least the following three things:

1. Greetings, asking about school/life
2. Complimenting & responses to compliment (on your friend's good personality trait—warm-heartedness & devotion)
3. Shifting to another topic (anything that you like to talk about; make sure that you have a smooth transition from the compliment to the new topic)

Appendix D
Naturalistic Role-Play Instructional Manual for Interlocutors

It is your task to lead the conversation in a flexible and natural way. Please read the description of the following two situations and identify yourself with the character "you" in it. If you have any questions, please feel free to ask. The whole role-play should **not exceed 10 minutes.**

Situation One:

You are enthusiastic about learning Chinese popular songs. One of your new friends, _________is a good singer and has won the first place in the university singing contest. You two know each other's names and have said hello to each other on occasion, but you have not yet had a chance to talk much.

It's now around 4pm and you are leaving school. You want to stop by a bookshop and have heard that there is one named "Foreign Language Bookstore" not far from school, but you do not know where it is. You are passing by the library and see your new classmate. You approach them to offer greetings. You talk while walking together. The talk should include, but may not be limited to the following points:

- *Please start the conversation by greeting your friend (your friend is listening to their MP4 player, so you need to shout loudly for them to notice you). Ask your friend about their classes. When it is most natural during the talk,* ***compliment*** *on their singing skills by referring to their winning in the recent singing contest.*
- *Please ask for directions to get to the Foreign Language Bookstore by bus.*
- *Please ask what time the bookstore is closed today.*
- *Please accept their company of taking the bus together to the bookstore. When it is most natural during the talk,* ***compliment*** *on their new MP4 player. You should pick a specific feature of the MP4 player to give your compliment.*
- *While you are both riding the bus, please talk about each other's hometown (e.g., compare the weather/ food/ population, etc.*
- *Before leaving the bookstore, please invite your new friend to a Karaoke party you organized in the Karaoke House next week. Your party will start at 6:00 p.m. next Friday, and you need to tell them the address and show them how to get there.*

Situation Two:

Today is the day of the Karaoke party. Now you are in the lobby of the Karaoke House chatting with friends and waiting for more to come, when you receive a call from _____________asking for directions. Your talk should include but may not be limited to the following points:

- *Mimic your cell-phone rings, pick it up. Listen to what your friend says first. (They will tell you about getting lost and ask for directions again).*

  Ask your friend where they are and tell them how to get to the Karaoke House.
- *(When your friend shows up at your doorstep) Greet them and invite them to come in. When it is most natural during the talk****, compliment*** *your friend on their appearance/clothing.*
- *Please respond to your friend's question expressing concern about your health (which is asked because they heard from their English teacher that you were sick).*
- *One of your colleagues brings their young daughter to the party. While the mother goes outside to pick up a phone call, the little girl starts to cry. Your friend goes up, talks and plays with the little girl until they stop crying. Please mimic the little girl crying (do it for fun!) and wait for your friend's "comforting". When it is most natural during the talk,* ***compliment*** *your friend on their being patient and loving towards the child.*
- *Ends the role-play by inviting your friend to sing a song.*

Appendix E
Conversational Instructional Manual for Interlocutors

Your main task is to be as sociable and friendly as possible while giving positive comments when you feel it is most natural. Be a good listener and attentive. Show concern to your friend. Find out as much information as possible about your friend. Please follow the instruction below. Your conversation should *not exceed 10 minutes.* Put a checkmark on each item after you have completed it.

1. (First do a brief introduction for each other if you haven't done so before). Start your conversation in a causal way by greeting the student at the very beginning, followed by *one* positive comment on the student's appearance/clothing.
2. Ask the student about their study/school work/life/future plans. When it is most natural during your talk, give *one* positive comment on the student's ability/work. (i.e., asking how school is going or how they did on a test/paper that was turned in recently or a presentation in another class or their future goals and what they are doing right now to achieve that goal: e.g., preparing to pass the GRE/TOEFL. An example compliment can be like the following: *It does sound like to me that you are an excellent/diligent student!* or *You must be really good at organizing your time. I can see that you are surely managing your work and play in a balanced way.* or *So are you planning on studying abroad? You surely have thought through all this! That's so impressive!*)
3. Ask the student to show you the object that they own and is proud to show and ask questions about it. When it is most natural during your talk, give *one* positive comment on this object. (e.g., *This does look like a really interesting book to me.* or *Is this a new model? It looks like a really cool cell phone to have.*)
4. While you're talking, try to identify *one* good personality trait of the student and compliment on it (e.g., perseverant/patient /humorous/responsible). Remember to keep chatting by asking follow-up questions based on the student's previous responses. If you are having a hard time identifying one specific personality trait to compliment on, you can ask the student to talk more about themselves to you. For example, you can ask probing questions such as "What is most important in your life? Why?" If the student came up with an answer such as "family/friends/career" You can ask further about their opinion on why this is viewed as the most important. After hearing the explanation, you can see the personality trait of the student and you can pay the compliment. Another way is to ask the student to share with you, one good deed they have done. When it is most natural during the conversation, give *one* positive comment on the student's personality traits (e.g., being warm-hearted/kind/patient, etc.)
5. End the conversation by thanking them for sharing with you.

Japanese EFL University Students and the Study Abroad Experience: Examining L2 Development and Program Satisfaction after Three Weeks in North America

Pino Cutrone*
Nagasaki University, Japan

Brien Robert Datzman
Nagasaki University, Japan

Abstract

This study is designed to shed light on the study abroad experiences of Japanese EFL university students with two goals in mind: to measure the effects of short-term study abroad on L2 proficiency in this context, and to provide EFL/ESL professionals with information that can help them develop practices to improve their study abroad programs. 98 university student participants were divided into five groups consisting of four experimental groups that studied abroad and one control group which stayed at home. Each group was given L2 proficiency tests at two points in time—prior to four of the groups embarking on three-week study abroad programs, and after the four groups returned from their time abroad. In addition, post-study abroad questionnaires and interviews were administered to assess students' attitudes and activities during the treatment period. The findings demonstrate that short-term study abroad did not have a great impact on students' TOEFL PBT scores and also provided detailed information to help study abroad professionals better understand the wants and needs of Japanese EFL university students. The authors discuss the implications of these findings and offer some suggestions for study abroad professionals and researchers to consider.

Keywords: study abroad, program development, second language acquisition, L2 proficiency

Introduction

"The real voyage of discovery is not in seeing new lands but in seeing with new eyes."
– Marcel Proust, French novelist –

As the profound and well-known quote by Marcel Proust touches upon, sojourns to new and unfamiliar lands can have a deep and lasting impact on a person's life. For university students, who are quite often only in the early stages of developing their identities and value systems, the experience of studying abroad can forever change and shape the way they view the world around them. While the value of the personal development aspect of study abroad (SA) is widely recognized and appreciated (O'Callagan, 2006), there seems to be some question as to what else students gain from their time abroad. In particular, there appears to be some debate surrounding the effects of SA on L2 (i.e., second language) proficiency. Until recently, there has existed a widespread belief that SA contexts, which provide natural settings and informal learning through out-of-class contact with the target language, lead to higher levels of proficiency than more formal in-class educational contexts where form-focused instruction is given (Tanaka & Ellis, 2003). However, in light of mounting evidence, this assumption may require a great deal of qualification.

* *Corresponding author. Tel. +81(0)95-819-2927; Email: k-pino@nagasaki-u.ac.jp; School of Global Humanities and Social Sciences, Nagasaki University, Nagasaki City, Nagasaki Prefecture, 852-8521, Japan*

As the next section will explore in greater detail, SA may affect gains in certain language-specific domains (such as fluency, pragmatics, and discursive abilities); however, it does not necessarily affect development in all aspects of learner proficiency (and most notably in the area of grammatical competence). Ironically, the grammatical domain of language development tends to be the main area that program administrators assess to determine whether students have benefited from SA (Collentine, 2009). This is particularly true in Japan where non-communicative examinations, and the washback cycle that comes with them, tend to define the curriculum (Cook, 2010, 2011). Concerning junior and senior high school contexts in Japan, much has been written about how the non-communicative nature of high stakes entrance examinations results in EFL instruction that stresses rote learning and grammar translation over fluency and communicative competence (Caine, 2005; Reesor, 2002; Sakui, 2004).

Similarly, though instructional methods will certainly vary according to institutional goals, a great number of tertiary institutions in Japan employ non-communicative tests such as the TOEFL PBT or the TOEIC to track student progress over time (Lee, Yoshizawa, & Shimabayashi, 2006; Vongpumivitch, 2013). Since these tests focus on listening, grammar, and reading, so, too, do the foci of many administrators and instructors in these settings. However, this is not to imply that the focus on these tests is misplaced and that they should not be important to Japanese people. To the contrary, obtaining high scores on these tests can be invaluable to Japanese people, as many Japanese companies still place a high value on TOEIC scores when hiring new employees (Takahashi, 2012), and TOEFL PBT scores can still be used to demonstrate a certain threshold in English proficiency (which differs among universities) to be able to study at many universities around the world (Hagerman, 2009). Rather, relating specifically to the scope of this paper, the writers question (and seek to examine more closely) whether such tests are really the most useful tools to assess the effects of SA on L2 proficiency, particularly in short-term programs.

One of the aims of this study is to help fill in the gap that appears to exist in the SA research literature. That is, various studies have looked at the gains learners make abroad in isolation (i.e., by simply looking at the pre- and post-SA abilities of one group); however, as Collentine (2009) points out, very few have actually compared learners' gains abroad with those of their student counterparts back home. Without a control group, it is difficult to know if the pre-post difference is a result of the intervention (SA) or not. Consequently, this study seeks to compare the pre- and post-TOEFL PBT scores of Japanese university EFL students who studied abroad (SA) with students from the same faculty who stayed at home (AH). Further, in addition to investigating the educational aspect of SA, this study also serves the practical managerial purpose of helping the writers improve upon the short-term SA packages that they offer to their students in the future. That is, by examining students' post-SA assessments, teachers will be able to gain a deeper understanding of the wants and needs of learners in this context and, thus, be able to make recommendations for future excursions.

Background to the Study

Regarding the Effects of SA on L2 Proficiency

In surveying the SA research literature, Coleman (1997) provides a list of twenty parameters for SA research, as follows: academic context, learning outcomes, age, program at home university, previous language learning, proficiency prior to departure, preparation, duration, outgoing/incoming group dynamics and structure, L1, L2, type of accommodation, who accommodation is shared with, program of non-language courses followed, who the program was taught by, which language the program was taught in, professional content, institutional support, and assessment. Dekeyser (1991) and Polanyi (1995) include personality and gender respectively in their lists of individual differences that can affect SA outcomes. Adhering to the central themes of this study, the review of the literature that follows will focus on two aspects of SA: the effects of SA on L2 proficiency and the organizational element of SA program development.

First, regarding research into the effects of SA on L2 proficiency, Carroll's (1967) original study of the relationship between the language proficiency of 2,782 American students majoring in French, German, Italian, and Russian and their SA experiences is often noted as the starting point. Carroll's study examined the L2 linguistic competence of college seniors, finding that even a short duration abroad resulted in higher levels of proficiency. Based on this solid claim of proof (and perhaps also influenced by anecdotal evidence and personal speculation), language professionals, school administrators, and students (and their parents) have traditionally assumed that SA contexts provide the best environment in which to acquire a foreign language. Over time however, it was realized that such early SA research lacked an overall systematic assessment of learners' gains, and concern with general proficiency shifted towards studies that focused on particular aspects of language competence (such as reading, writing, listening, and speaking). By narrowing the focus of each investigation, incorporating more diverse methods, and adopting a wide variety of theoretical approaches, researchers began to shed new light on language learning in SA contexts. First, to the surprise of many, empirical studies consistently supported the notion that SA contexts did not necessarily bring about higher proficiency; educational contexts often predicted higher proficiency, particularly in the area of grammatical competence. Second, in SA contexts, the amount of contact with the L2 seemed to have less influence on language learning than the type of contact, which differed depending on learners' initial levels of proficiency.

Combining the surveys of the research literature in this area conducted by Freed (1990, 1993, 1995, 1998) and Coleman (1997), Tanaka and Ellis (2003) provide us with the following summary concerning the effects of SA on language learning:

1. Accuracy and complexity, measured in terms of frequency of mistakes, sentence length or syntactic complexity in oral production, did not change in any noticeable way.
2. Gains in fluency, in terms of the speaking rate (syllables per minute) or phonation/time ratio (percentage of total time spent speaking), were strong.
3. Overall oral proficiency scores, measured by the ACTFL Oral Proficiency Interview (OPI), were higher in learners in study-abroad programs than in learners who did not participate.
4. Gains in overall oral proficiency scores were stronger than gains in test scores on grammar, listening, and reading.
5. Vocabulary gains, measured by vocabulary tests, were stronger than those of comparable students who did not participate in a study-abroad program.
6. The higher the students' initial level of proficiency, the lower the gains in proficiency as a result of studying abroad. (p. 67)

To summarize, much of the research to date seems to suggest that, overall, SA does not greatly impact grammar, listening, and reading (i.e., in terms of accuracy and complexity of language use), particularly in higher-level students. Rather, an increase in natural exposure to the target language in a SA context appears to contribute more to fluency and naturalness of speech (i.e., higher speech rate, and fewer disfluency features such as false starts, repetitions, and corrections). Accordingly, the findings presented above would seem to challenge some of the initial (and, in many cases, current) assumptions of the effects of SA on language learning.

Moreover, concerning levels of proficiency specifically, several studies have demonstrated the great impact that SA can have on learners with lower proficiency levels. With this in mind however, the research in this particular area is in its infancy and a great deal of work remains to be done. In terms that have yet to be precisely defined, researchers have begun to speculate that there exists a certain threshold that learners must reach to benefit fully from the SA context of learning (Lafford & Collentine, 2006). According to Collentine (2009), "there are most likely specific domains that require a particular developmental threshold for overall gains to occur" (p. 221).

In addition, another area of SA research that requires more attention is duration, i.e., how varying lengths of time abroad affect SA-related outcomes. Presently, it is unclear what an ideal duration of a SA program might be, as results published in this area thus far are often conflicting and inconclusive, reflecting insufficient sample sizes, the heterogeneity of the studied populations, and the idiosyncratic nature of the SA experience. One of the areas in which researchers appear to be reaching a consensus is concerning grammatical development, i.e., gains in grammatical competence made by SA students are not thought to outpace those made by AH students, at least within the timeframe of a semester to a year abroad (Collentine, 2004, 2009; DeKeyser, 1990, 1991).

Furthermore, in broader terms, the conventional wisdom in the field of SA that longer is better is supported by the findings of Dwyer (2004). In their landmark study, Dwyer surveyed 3,723 former SA students in the US from the previous 50 years. In measuring the longitudinal effects of SA on academic, career, intercultural, and personal development, Dwyer (2004) found that "study abroad has a significant impact on students in the areas of continued language use, academic attainment measures, intercultural and personal development, and career choices" (p. 161). While various researchers have questioned the impact of short-term SA programs on student motivation (Freed, 1990; Sasaki, 2011), there is evidence demonstrating the benefits of short-term SA programs in various language domains. For instance, Llanes and Muñoz (2009) report on how 24 Spanish students of English were able to improve their oral fluency after 3-4 weeks studying abroad in an English-speaking country. In another study, Evans and Fisher (2005) tracked the development of 68 British pupils after only 6-11 days studying abroad in France and found dramatic and sustained improvements in their L2 listening and writing skills.

Finally, it is important to keep in mind that individual differences can have a great impact on SA-related outcomes. In fact, as Huebner (1995) has pointed out, SA tends to accentuate individual differences, as "the overseas experience seems to result in a much wider variety of performances and behaviors than does study at home" (p. 191). This was certainly the case in studies conducted by Brecht and Davidson (1991) and Brecht, Davidson, and Ginsberg (1995) who examined 668 American learners' acquisition of Russian in a SA context and found a great deal of individual variation in students' attitudes and L2 performance in this context. Undoubtedly, second language acquisition can become quite complicated by the socio-cognitive and socio-cultural pressures that learners face in the SA context, which Collentine (2009) describes as "a situation that sends many more messages to learners than does the AH context as to the complete repertoire of skills and behaviors one needs to be communicatively functional" (p. 226).

The Development of SA Programs for JEFLs

In addition to investigating the educational aspect of SA, this paper examines the administrative and organizational aspects of running SA programs for Japanese EFL learners. First of all, it should be noted that SA in this context is currently facing a time of crisis, as Japanese EFL students are not studying abroad like they used to. After peaking in 2004 with 82,945 students, the number of Japanese studying abroad has continued to decline every year, dropping to a mere 58,060 in 2010 (Nagata, 2013). This 30 % decrease in less than 10 years has largely been attributed to the shrinking number of students and the fact that they perhaps tend to be more inward looking than students in the past. In light of this recent trend, the Japanese government, fully recognizing the benefits of the SA experience, has initiated several programs and incentives, which included doubling the number of SA scholarships available in the 2014 academic year, to encourage more Japanese university students to pursue SA (Torres, 2013). Accordingly, in this state of flux, it is of increasing importance that SA administrators in Japan work to understand the wants and needs of their learners and to provide them with the best SA experiences possible.

Given that many SA administrators serve primarily as academic staff in Japanese educational institutions, they often have larger programmatic goals in mind and may be especially concerned with the results of department-driven test scores (Hidasi, 2004). Thus, it is not surprising that few studies have been published concerning student attitudes and satisfaction relative to the SA experience. Nonetheless, even within this limited body of work, the positive effect of SA on Japanese EFL learners' affective dimensions is evident. For instance, in a study involving 24 Japanese university students on a six-week summer study program in the US, Geis and Fukushima (1997) observed increased levels of motivation in the students' classroom behavior upon returning to Japan and resuming their studies. In another study, which involved 139 Japanese high school students on a one-year study program in the US, Yashima (1999) found that students who tended to be more extroverted and less

Japanese-centered were better able to adjust to their SA environment. Similarly, in a study involving 60 Japanese high school students on a SA program in the US, Yashima, Zenuk-Nishide, and Shimizu (2004) examined the effect of willingness to communicate (WTC) on students' perceived satisfaction of their SA experience through a questionnaire and found that pre-program WTC and self-confidence led to greater levels of satisfaction and opportunities to communicate.

While various positive aspects of student attitudes and satisfaction have been duly noted, two fairly recent studies offer some insights into how administrators can improve their students' SA experiences. First, based on qualitative and quantitative data collected from 28 Japanese students (ranging from 19-25 years old) returning from a short-term SA program at an Australian university, Furmanovsky (2005) reveals some of the participants' perceived weaknesses, which includes the inability to participate fully in class discussions, weak presentation skills, the inability to explain aspects of Japanese culture in English, and a poor grasp of contemporary world affairs. Furthermore, the participants in Furmanovsky's (2005) study expressed "a clear preference for smaller classes, in which relaxed discussion or an exchange of views with other students within groups is possible" (p. 9).

Second, another study that may help shed light on SA program reform was conducted by McIntyre (2007), who sought to determine a set of needs and goals that could be used to develop a preparation program for long-term overseas study by Japanese EFL university students. Based on interview data collected from five Japanese university students upon returning from a SA program in the US, McIntyre found that the needs and goals of students could be grouped into three main areas: cognitive/academic needs, linguistic needs, and socio-cultural needs. Educationally, participants placed a high premium on academic reading, writing, presentation, and discussion skills. From a personal perspective, participants expressed how important the personal relationships they develop while abroad are to their overall SA experience. In particular, they valued being attended to and taken care of by their host-culture teachers and were especially concerned with their ability to forge long-lasting friendships with members of the host culture.

Research Questions

Research into the area of SA is multifaceted, and this study, which seeks to help determine the effects of short-term SA on students' L2 proficiency and attitudes, is designed with the following objectives in mind. First, by drawing on the research to date, this study reassesses the general assumption that SA leads to higher L2 proficiency levels, and, more specifically, reexamines whether standardized tests, such as the TOEFL, are adequate measurements to reflect success in (and from) SA. Second, by examining students' post-SA assessments, this study aims to provide EFL/ESL professionals with information that can help them develop practices to improve their SA programs, and, more specifically, to offer insights designed to help them better understand the wants and needs of Japanese EFL university students. To shed light on these areas of SA, the following research questions (RQs) were formulated: (RQ 1) What impact did the SA experience have on students' TOEFL PBT scores? and (RQ 2) How did students feel about their study abroad experiences?

Methodology

Design

This study examines the L2 proficiency and attitudes of five groups of Japanese EFL university students, comprised of four experimental groups that studied abroad (in four separate SA programs that lasted approximately three weeks) and one control group that stayed at home. Each group was given TOEFL PBT tests at two points in time: prior to the four groups going abroad, and after they returned from their time abroad. Additionally, post-SA questionnaires and interviews were conducted to determine how the SA students felt about their time abroad as well as what type of activities the AH students engaged in while their counterparts studied abroad.

Participants

The 98 participants were all freshmen students at a national university in Japan (71 females and 27 males), who were enrolled in a faculty that focuses on the study of global humanities and social sciences and that emphasizes the study of English. When this study began, participants were on average at an intermediate level of English proficiency (as reflected by their TOEFL PBT scores), between 18 and 20 years old, and had studied English for 6.5 years on average (including a collective six years in junior and senior high school). The students had been enrolled at the university for six months and were all taking courses related to English study. Participating of their own free will and understanding the nature of the study, all participants were given explicit instructions (i.e., verbal and written, in both English and Japanese) regarding this study and their role in it.

As all of the participants of this study were comprised of students taught by the researchers, the Japanese EFL learners in this study constituted an opportunistic sample. That is, all first-year students in the above mentioned faculty (who happened to be accessible to the researchers) were approached to participate in the study. The researchers had no control in choosing which students studied abroad and which ones stayed at home at the time of this study, as such decisions were made by the students themselves within the parameters of the SA options made available to them by their faculty and university. Most, if not all, of the AH students chose not to study abroad due to financial restraints and/or because they were hoping that they would have other opportunities in the future to study abroad in destinations that were more desirable to them.

Study-Abroad Contexts

As shown in Table 1 below, this study comprises five groups: four experimental groups that studied abroad and one control group that stayed at home. The four SA programs involved in this study shared several similarities: they were all based at major universities in North America, and the ESL classroom instruction each program provided focused on the development of the four major skills of language competence (i.e., reading, writing, listening, and speaking). In addition, each program provided a range of socio-cultural (SC) activities, which included educational field trips to museums and historical and cultural landmarks, sightseeing excursions, group shopping trips, parties with teachers, local students, and other international students, etc. The duration of each program was also similar in that Groups A, B, and D studied abroad for three weeks, while Group C studied abroad for 3.5 weeks.

In some ways, the programs contained some important differences. First, the groups who attended the Canadian universities (Groups A, B, and C) stayed with host families, and the average number of hours of ESL instruction each group received each week was 24. In comparison, the group that attended an American university (Group D) stayed in a student dormitory, which they shared predominantly with their classmates from Japan, and received an average of 18 hours of ESL instruction per week. Of the three groups that stayed with host families (Groups A, B, and C), members of Groups of A and B were often grouped together with one or more international students. In several cases these roommates were Japanese students from other universities, and a few students in Group B were paired together with their classmates from the same university in Japan. Further, the make-up of class members in each program was somewhat different. Whereas members of Groups A and B were put in classes that were mixed with students from other countries to varying degrees, Groups C and D were in customized classes that contained only Japanese students from the home institution.

At-Home Context

Concerning the control group used in this study, members of Group E, who stayed at home (AH) did not receive any formal EFL instruction while the other students studied abroad. However, it is important to note that after the first TOEFL test (i.e., the pre-test that all participants took), all participants in this study (i.e., experimental and control groups) received similar instruction in the final month and a half of the semester before the SA students departed for North America. All students were enrolled in classes consisting of General English, English Communication, Reading and Discussion, Reading and Writing, English Pronunciation and Phonetics, and First-Year Seminar.

Table 1
The Five Groups Used in this Study

Groups (* Control)	Number in each Group	Location	Length	Hours of Study / Week	Content of Study	Make-up of Class Members	Accomodation Settings (alone or with other students; if with others, monolingual or mixed nationalities)
A	13	Central Canada	3 weeks	24	4 skills (+ SC acitivities)	partially mixed nationalities	Homestay (with other students; partially mixed nationalities)
B	16	Central Canada	3 weeks	24	4 skills (+ SC acitivities)	mixed nationalities	Homestay (with other students; partially mixed nationalities)
C	13	Western Canada	3.5 weeks	24	4 skills (+ SC acitivities)	monolingual (all Japanese)	Homestay (mostly alone with host family)
D	15	Western United States	3 weeks	18	4 skills (+ SC acitivities)	monolingual (all Japanese)	Dormitory (with other students; monolingual - all Japanese)
* E	41	Japan	N/A	Variable	Variable	independent	N/A

Instruments

Three instruments were used in this study: questionnaires, interviews, and language proficiency tests. The two questionnaires used in this study were designed respectively (1) to assess the efficacy of each SA program in order to make improvements for future excursions and (2) to monitor the activities of the AH students while their SA counterparts were abroad. First, the post-SA questionnaire for SA students consisted of ten questions, which comprised a 27-item inventory (see Appendix A). Seven questions were open-ended and were general in nature (with the exception of Question 1 which was a more focused open-ended question, as it asked specifically for identifying information). The remaining three questions were comprised of 20 items that were divided into three categories of assessment (i.e., regarding academic experience, homestay experience, and cultural experience). The three categories comprised 12, 10, and 8 closed questions respectively, with statements on a Likert-scale ranging in perceptions from *poor* to *excellent* or ranging in degree from *a little* to *a great deal*. Second, the questionnaire for AH students consisted of five items, three of which were closed questions and two of which were open-ended questions (see Appendix B). Both questionnaires were administered in English; however, Japanese support (and translation) was provided to participants who requested clarification. The questionnaire was piloted on a group of five Japanese EFL university students (who did not participate in this study), which resulted in the researchers modifying the wording of two of the items on the post-SA questionnaire.

Another data collection instrument used in this study was interviews. Each interview was conducted in the primary researcher's office, with only the interviewee (i.e., study participant) and the interviewer (i.e., the primary researcher) present at the time of the interview. The aims of the interview were twofold: to follow up on any ambiguities and/or noteworthy responses from the participant's questionnaire, and to dig deeper into how students felt about each aspect of their SA experience. The interviews were semi-structured in that the interviewer had a general plan for the interviews, but did not enter with a predetermined set of questions, as some of the questions were guided by the context of the conversations and the responses of the interviewee. The interviews were conducted in English; however, Japanese support was provided to participants who requested clarification.

Lastly, in order to measure overall English proficiency, a paper-based version of the Test of English as a Foreign Language (TOEFL) was administered. Specifically, the paper-based version of the TOEFL used in this study refers to the TOEFL Institutional Testing Program (ITP), which is an internal testing program used in the participants' home university to monitor students' English progress over time. The TOEFL ITP Level 1 adheres to the same academic standard, composition, and scoring system of the regular TOEFL Paper-Based Test (PBT).

All questions are multiple choice and students answer questions by filling in an answer sheet. The tests evaluate skills in the following three areas:

- Section 1 (Listening Comprehension) measures the ability to understand spoken English as it is used in colleges and universities,
- Section 2 (Grammar: Structure and Written Expression) measures recognition of selected structural and grammatical points in standard written English, and
- Section 3 (Reading Comprehension) measures the ability to read and understand academic reading material in English.

Data Collection Procedures

Table 2 shows the schedule for the administration of the pre- and post-SA TOEFL tests, and the post-SA satisfaction questionnaires and interviews. The pre-SA TOEFL was administered some two months before the majority of the students began the SA program, while the post-SA TOEFL took place two weeks after the end of the SA program. The post-SA satisfaction questionnaire and the AH questionnaire were administered approximately one week after students returned, and the post-SA interviews were conducted approximately one to two weeks after students returned.

Table 2
Schedule of the Study

Instrument	Time Administered
TOEFL 1	2 months before study abroad
TOEFL 2	2 weeks after study abroad
Post-SA Satisfaction and AH Questionnaires	1 week after study abroad
Interviews	1-2 weeks after study abroad

Data Analysis

This section describes how the data produced by the tests, questionnaires, and interviews were analyzed in this study. First, concerning the TOEFL tests, paired samples t-tests were used to determine whether the differences between the means of pre-SA and post-SA TOEFL scores were statistically significant for each group. The Statistical Package for the Social Sciences (SPSS), version 21.0, was used to analyse such data in this study. Two-tailed tests were used, and alpha levels (α) were set at 0.05. However, as the groups in this study were relatively small, probability statistics have to be viewed with caution. Thus, considering the possibility of Type 1 errors (i.e., the false rejection of the null hypothesis) occurring, probabilities less than the more stringent 0.01 and 0.001 levels will also be shown and discussed.

Second, the post-SA questionnaires and interviews were designed to shed light on how participants from each group felt about their SA experience as well as to discover what types of activities the AH students engaged in while their SA counterparts studied abroad. Responses to the closed questions on the questionnaire produced quantitative data and will, thus, be presented using descriptive statistics that denote the general characteristics of the sample. Responses to the open-ended questions on the questionnaire and interview produced qualitative data, which will be presented in two ways: based on key aspects of SA that the researchers have chosen for further inquiry and exploration, and other previously untargeted themes that emerged throughout the course of this study (i.e., which often came to light when the interviewer delved deeper into particular student responses and/or certain trends that began to emerge in the data).

Results

TOEFL Scores

This section has been divided into three parts. The first part will report the results concerning student TOEFL scores before and after SA, while the second and third parts will present the quantitative and qualitative aspects o f students' post-SA assessments respectively. Corresponding to each of the four experimental groups (A, B, C, and D) respectively, Tables 3, 4, 5, and 6 show the results of pre- and post-SA TOEFL scores. It should be noted however that TOEFL scores were only available for 71 of the 98 participants, as many of the participants were absent on one (or both) of the testing days.

As shown below, the four experimental groups exhibited similar improvements in TOEFL scores over time. The mean total TOEFL scores for Groups A, B, C, and D improved by approximately 22, 20, 18, and 21 points respectively. The difference between pre-test and post-test scores was statistically significant for all groups, which reflected an improvement of 2.6, 4.1, 3.7, and 4.4 % respectively. As the breakdown of the sections in each table demonstrates, the largest gain in proficiency was seen in the Reading Comprehension Section of the TOEFL for all groups (10.7, 11, 6.5, and 6.2 % respectively) and the smallest in the Grammar Section (0, 1, .9, and 1 % respectively).

Table 7 presents the TOEFL score results of the experimental groups (Groups A-D: all SA students) together, while Table 8 reports on the results of the control group (Group E: the AH students). Combining the findings shown in Tables 7 and 8, Figure 1 collectively compares the pre- and post-SA TOEFL PBT scores of SA students with those of AH students. As Figure 1 illustrates, the students who did not go abroad exhibited greater improvements in TOEFL scores overall; however, the scores of both groups generally followed the same path. The mean total TOEFL score for SA students improved by approximately 20.2 points, while that of AH students improved by 28 points. The difference between pre-test and post-test scores was statistically significant for both groups, which reflected an improvement of 4.2 and 5.8 % respectively. As the breakdown of the sections in Tables 7 and 8 demonstrate, the largest gain in proficiency was observed in the Reading Comprehension Section of the TOEFL for both groups (8.6 and 10.7 % respectively) and the smallest in the Grammar Section (1.1 and 3.2 % respectively).

Table 3
Group A's TOEFL Scores Before and After Studying Abroad

N = 9	Pre-test Mean (SD)	Post-test Mean (SD)	Mean Diff.	(%)	t	p
Listening	48.44 (3.21)	49 (3.24)	.56	1.16	-.679	.516
Grammar	46.89 (1.76)	46.89 (2.15)	0	0	0	1
Reading	47 (1.73)	52 (3.71)	5	10.64	-4.29	.003**
Total	470.78 (19.55)	492.78 (24.23)	22	2.55	-4.08	.004**

(Mean difference of Pre-test → Post-test significant at p<.05 level = *; significant at p<.01 level =**, and significant at p<.01 level =***)

Table 4
Group B's TOEFL Scores Before and After Studying Abroad

N = 13	Pre-test Mean (SD)	Post-test Mean (SD)	Mean Diff.	(%)	t	p
Listening	51.92 (3.38)	51.77 (3)	.15	.29	.273	.790
Grammar	48.62 (3.53)	49.62 (4.21)	1	2.06	-.952	.360
Reading	48.23 (2.49)	53.54 (5.03)	5.31	11.01	-3.28	.007**
Total	496 (18.44)	516.39 (18.51)	20.39	4.11	-3.65	.003**

(Mean difference of Pre-test → Post-test significant at p<.05 level = *; significant at p<.01 level =**, and significant at p<.01 level =***)

Table 5
Group C's TOEFL Scores Before and After Studying Abroad

N = 12	Pre-test Mean (SD)	Post-test Mean (SD)	Mean Diff.	(%)	t	p
Listening	48.75 (3.91)	50.83 (2.62)	2.08	4.27	-2.29	.043*
Grammar	48.42 (3.94)	48.5 (4.52)	.08	.17	-.05	.961
Reading	47.42 (3.83)	50.5 (3)	3.08	6.5	-3.6	.004**
Total	481.92 (29.91)	499.50 (21.76)	17.58	3.65	-2.21	.049*

*(Mean difference of Pre-test → Post-test significant at p<.05 level = *; significant at p<.01 level =**, and significant at p<.001 level =***)*

Table 6
Group D's TOEFL Scores Before and After Studying Abroad

N = 10	Pre-test Mean (SD)	Post-test Mean (SD)	Mean Diff.	(%)	t	p
Listening	50.3 (3.37)	51.9 (3.47)	1.6	3.18	-3.21	.011*
Grammar	49 (4.45)	50 (3.74)	1	2.04	-.72	.488
Reading	48.2 (5.07)	51.2 (4.05)	3	6.22	-2.5	.034*
Total	491.8 (35.19)	513.2 (30.97)	21.4	4.35	-4.44	.002**

(Mean difference of Pre-test → Post-test significant at p<.05 level = *; significant at p<.01 level =**, and significant at p<.001 level =***)

Table 7
Pre- and Post-SA TOEFL Scores of all SA Students

N = 44	Pre-test Mean (SD)	Post-test Mean (SD)	Mean Diff.	(%)	t	P
Listening	49.98 (3.7)	50.98 (3.15)	1	2	-2.67	.011*
Grammar	48.3 (3.57)	48.84 (3.91)	.55	1.12	-.857	.396
Reading	47.75 (3.42)	51.86 (4.1)	4.11	8.61	-6.39	.000***
Total	486.05 (27.3)	506.23 (24.9)	20.18	4.15	-6.63	.000***

(Mean difference of Pre-test → Post-test significant at p<.05 level = *; significant at p<.01 level =**, and significant at p<.001 level =***)

Table 8
The TOEFL Scores of AH Students at Two Points in Time

N = 27	Pre-test Mean (SD)	Post-test Mean (SD)	Mean Diff.	(%)	t	p
Listening	49.89 (4.47)	51.37 (4.86)	1.48	2.97	-2.3	.03*
Grammar	47.74 (3.47)	49.25 (5.04)	1.52	3.16	-1.9	.068
Reading	46.19 (5.55)	51.15 (4.31)	4.96	10.74	-4.22	.000***
Total	479.33 (35.51)	507.3 (37.88)	27.96	5.84	-4.86	.000***

(Mean difference of Pre-test → Post-test significant at p<.05 level = *; significant at p<.01 level =**, and significant at p<.001 level =***)

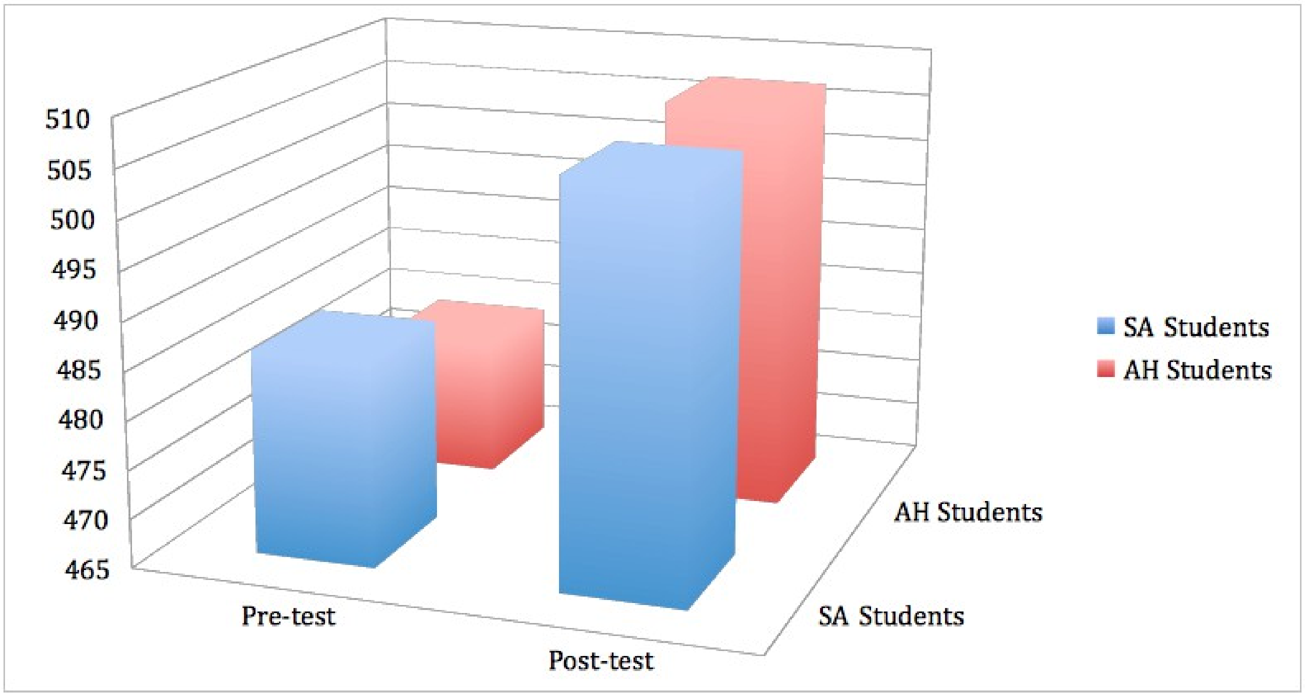

Figure 1. Comparing TOEFL Scores of SA and AH Students at Two Points in Time

Lastly, concerning the control group (Group E), 28 of 41 AH students responded to the questionnaire designed to monitor the activities of the students who did not go abroad. Although the AH students did not receive any formal EFL instruction during this time, several of them engaged in independent study to varying degrees: while no student reported to using English *a great deal*, 2 students responded to using English *much*, 9 students reported to using it *somewhat*, 12 said that they used it *a little*, and 5 answered *not at all*. Some of the activities students took part in were watching English movies, communicating with foreign friends, reading graded readers, and studying for the TOEFL test. Regarding how much time students spent studying for the TOEFL test specifically, 4 students reported *somewhat*, 14 responded *a little*, and 10 said *not at all*. The main activities students mentioned were studying from TOEFL textbooks and attempting practice tests.

Quantitative Data Reporting on Students' Post-SA Assessments

Tables 9, 10, and 11 report on the following three facets of students' post-assessment questionnaires respectively: the academic aspect of SA, information pertaining to student accommodations, and the cultural experience. Each of these tables presents the average score for each group (A, B, C, and D) corresponding to the items on the questionnaire that were closed questions. Therefore, corresponding to ratings on a Likert-scale ranging in perceptions from *poor* (1) to *excellent* (5) or in degree from *a little* (1) to *a great deal* (5), larger scores in the tables below would convey a greater sense of satisfaction or achievement. As these tables demonstrate, ratings across groups, as well as across categories, were generally quite high overall; however, there were a few notable exceptions in which groups expressed less satisfaction (i.e., noted when a group's mean rating on a particular item was ≤ 3). For instance, as shown in Table 9 (concerning the Academic aspect of SA), Group D's mean rating of 2.8 on Item B suggests that they may have not been wholly satisfied with the preparation they received from their home university pre-departure. In comparison, Groups A, B, and C's mean ratings were 3.46, 3.5, and 3.31 respectively.

Moreover, as shown in Table 10 (concerning students' SA accommodations), the mean ratings for Group D (who, unlike the other groups, stayed in a student dormitory) were below 3 on 8 of 10 items and consistently lower than the other three groups on all 10 items. Finally, as shown in Table 11 (concerning the cultural aspect of SA), Group C and D's mean ratings of 2.92 and 3 on Item H indicates that the members of these groups may have not been wholly satisfied with the degree to which they were able to make friends with people in the local community off-campus. Following this same theme, it appears that members in all groups may not have been wholly satisfied with the amount of contact they had with locals off-campus, as demonstrated by the fact that the mean ratings were ≤ 3.5 for all groups on all items (F, G, and H) concerning contact with local people. The issues and themes brought to light by these results warrant further inquiry and explanation and were explored in the follow-up interviews.

Table 9
Students' Post-SA Assessments Concerning the Academic Aspects of SA

Groups	A N = 13	B N = 16	C N = 13	D N = 15	Total Avg.
A. The preparation you received from your home university pre-departure	3.46	3.5	3.31	2.8	3.27
B. The orientation and support you received from the host university upon arrival	4.15	3.94	3.92	3.33	3.84
C. The accessibility and helpfulness of the host university's on-site director	4	4.5	4	3.87	4.09
D. The degree to which the study abroad program met your expectations overall	3.62	4.56	3.77	3.73	3.92
E. The safety of the host university's campus	4.77	4.75	4.69	4.60	4.70
F. The overall classroom instruction you received at the host university	4.15	3.94	3.92	4.57	4.15
G. Your instructors' course preparation and delivery at the host university	4.08	4.38	4.08	4.47	4.25
H. The reasonableness of assignments at the host university	3.92	4.06	3.85	3.93	3.94
I. The usefulness of assignments at the host university	3.85	4.31	4.15	4.27	4.15
J. The degree to which classroom instruction at the host university matched your English proficiency level	3.77	4	3.54	3.93	3.81
K. The degree to which you think your English communication skills improved overall	3.85	4.33	3.92	3.67	3.94
L. The degree to which you think your TOEFL score improved	3.08	3.19	3.08	3	3.09

Table 10
Students' Post-SA Assessments Concerning their SA Accomodation

Groups	A N = 13	B N = 16	C N = 13	D N = 15	Total Avg.
A. The degree to which the homestay/dormitory experience met your expectations	4.15	4.25	4	2.69	3.77
B. The cleanliness of the accommodation you stayed in	4.38	4.31	4.23	3.62	4.14
C. The friendliness of your host family (or other inhabitants of the dormitory)	4.38	4.63	4.23	3.23	4.12
D. The food your host family (or dormitory food services) prepared for you	4.15	4.31	4.15	3	3.90
E. The safety of the home/dormitory and area that you stayed in	4.46	4.81	4.46	3.85	4.4
F. The support you received from your host family/dormitory regarding your commute to and from your host university	4.23	4.25	4.15	3.62	4.06
G. The degree to which your host family (or other inhabitants of the dormitory) was interested in Japan and Japanese culture	4.15	4.38	3.69	2.69	3.73
H. The degree to which you were able to communicate in English with your host family (or other inhabitants of the dormitory)	3.92	4.25	4.15	2.69	3.75
I. The degree of confidence that you will keep in touch with your host family (or other inhabitants of the dormitory) in the future	3.77	4.06	3.69	2.46	3.5
J. The degree of support you received from your host university regarding any issues you had with your accommodations	4	4.11	4.13	3	3.81

Table 11
Students' Post-SA Assessments Concerning the Cultural Aspects of SA

Groups	A N = 13	B N = 16	C N = 13	D N = 15	Total Avg.
A. The degree to which the overall cultural experience met your expectations	4.23	4.38	4.5	4.07	4.3
B. The sociocultural activities you participated in, as organized by your host university	4.23	4.5	4	3.87	4.15
C. The sociocultural activities you participated in, as organized independently and informally by you and/or your friends	4.31	4.19	4.15	3.64	4.07
D. The sociocultural activities you participated in, as organized by your host family	4.08	3.75	3.69	3.08	3.65
E. The degree to which you were able to learn about and experience the host culture	4.23	4.06	4.46	3.36	4.03
F. The amount of contact you had with people in the local community off-campus	3.31	3.63	3.31	3.29	3.39
G. The degree to which you were able to communicate successfully in English with people in the local community off-campus	3.46	3.5	3.46	3.21	3.41
H. The degree to which you were able to make friends with people in the local community off-campus	3.23	3.38	2.92	3	3.13

Qualitative Data Reporting on Students' Post-SA Assessments

In general, the quantifiable aspects of students' post-SA assessments were borne out in the qualitative data. Due to space limitations however, it is not possible to include all responses from students who went abroad. Therefore, to represent the various patterns found in student responses, and in answer to questions pertaining to key concepts in this study (i.e., namely, RQ 2), the following general summarizations are provided:

- Most, if not all, of the participants indicated that their SA experience was an overwhelmingly positive one.
- Many of the participants admitted to feeling anxious at first about going abroad.
- Some of the participants acknowledged that their fears about going abroad could have been allayed somewhat if they had received more information (and much earlier) prior to departure.
- Many participants linked post-SA satisfaction with how much individual attention they received in their classes and in their homestay/dormitory contexts.
- Specifically, the perceived quality of participants' SA experience seemed to be largely dependent on personal relationships, and, in particular, on the friendships they formed with their teachers, host families, and other international students.
- Participants tended to prefer staying with a host family to staying in a student dormitory.
- Most participants preferred not to share a host family with another student and were particularly disappointed when another Japanese student was placed with them.
- Most participants preferred their classes to contain as few Japanese students as possible.
- In general, participants preferred small and intimate classes where they had more opportunities to speak.
- Most, if not all, participants preferred socio-cultural activities and coursework that related to the

practical usage of English over coursework that was more academic in nature (such as Academic Reading, Writing, and TOEFL/TOEIC preparatory courses).

- Generally speaking, participants preferred classes taught by native speakers of English.
- Many of the participants felt that the length of their SA program was too short and expressed the desire to have it extended by another week or two.
- Several participants expressed disappointment at the abrupt end of their SA program, i.e., upon returning from their SA program, they felt they had little opportunity to use English and/or no forum to follow up on (and discuss) their SA experiences.

Discussion

The good news for students in this context is that they were able to significantly improve their TOEFL PBT scores in a relatively short period of time; however, in answer to RQ 1, SA appears to have had very little impact on student TOEFL PBT scores. In fact, as shown by the skills measured on the TOEFL PBT, AH students actually showed greater improvements than SA students (28 points vis-à-vis 20 points). This was not surprising as members of the AH group were not distracted by the many communicative demands that members of the SA group had to face, not to mention the fact that some members of the AH group actually dedicated specific attention to the TOEFL test in their independent studies. However, the fact that the standard deviation was quite high for the AH group would seem to suggest considerable variability with the performances within this group.

While the general findings of this study are consistent with those of recent studies in this area, it was somewhat interesting to see all groups in this study significantly improve their TOEFL PBT scores. There are several possible reasons to explain this. First, the increase in scores from the pre-test to the post-test may simply be attributed to the experience factor, i.e., the fact that students taking the post-test were much better prepared (and far less anxious) because they knew what to expect from their previous experience with the pre-test (which was the first time most of the students had ever taken the TOEFL). Second, as shown by the marked improvement in the Reading Section across the board ($p<.001$ level), improvements in this area may have had more to do with the home institution's curriculum, as pre-SA instruction seemed to emphasize the development of reading skills.

Moreover, while the final part of the Results Section would seem to provide a general answer to RQ 2, some of the summarizations made therein would seem to require a degree of qualification. In particular, a major theme in student responses was that they wanted as much individual attention as possible. This speaks to the lack of L2 practice opportunities afforded Japanese EFL learners in a largely homogeneous nation such as Japan. Simply put, the number one fear among many Japanese students taking part in SA programs seems to be that they will waste their time abroad speaking Japanese. Thus, it makes sense for SA professionals to avoid putting Japanese students in monolingual settings whenever possible and to make a concerted effort to provide such learners with practical contexts for authentic L2 use.

With these goals in mind however, active participation in the classroom may in some cases be difficult to attain, as the research is littered with studies and anecdotes about how shy and passive Japanese EFL learners supposedly are (Greer, 2000; Townsend & Danling, 1998). Researchers such as Anderson (1993) have disputed these types of characterizations, claiming them to be mere reflections of Western ethnocentrism. That is, Japanese learners actually do speak, and sometimes they speak a lot, but the contexts in which they communicate are culturally sanctioned and do not always correspond to the cultural codes of the West. Adhering to the well-known Japanese proverb *the nail that sticks up gets hammered down*, many Japanese learners tend to be reluctant to speak in contexts where they will stand out in front of their peers. Thus, it is imperative that educators in SA contexts be aware of (and sensitive to) the social and cultural codes for language use that do indeed exist. Further, as one of the writers noted previously, one instructional approach that may help draw out Japanese EFL learners involves a combination of techniques that draw on the dynamics of the Japanese classroom, with strategies that promote a Western style of interaction (Cutrone, 2009).

Conclusion

Before any conclusions are presented, it is necessary to consider the limitations of this study. First, having focused mainly on the SA-AH distinction, it is important to point out that there exist several phenomena not examined in this study that may affect student performances in certain contexts. Specifically, an individual's behaviour at any point in time will also be greatly influenced by numerous variables such as personality , mood, age, gender, class, the context of the situation, group dynamics, etc. As the section below mentions, more research concerning individual differences in SA contexts may be beneficial. Second, concerning research design, the aspect that dealt with students' L2 performances involved mainly quantitative data and analysis. Undoubtedly, additional insights into the impacts of SA could have been gained by probing a bit more, with a qualitative approach, into the participants' perceived gains in, or self-assessments of, aspects of their language use. Presently, the authors are embarking on a larger scale study to build on the findings of this exploratory study.

This study has explored two important aspects of the SA experience: (1) the effects of short-term SA on L2 proficiency and (2) SA program development in the Japanese university EFL context. Combining the findings of this study with those of previous studies, the following conclusions and suggestions are provided:

1. Regarding the Effects of SA on L2 Proficiency:
   - Widely-used tests such as TOEIC and TOEFL PBT do not seem to be adequate measures for success in (or from) short-term SA.
   - Readily available tests that include Oral Proficiency Interviews (OPI) and focus on fluency in spoken discourse, such as those administered by ACTFL and (to a lesser degree) IELTS, seem to be more well-suited to measuring the effects of SA (i.e., even the TOEFL iBT which has a speaking component is not ideal, as it focuses on students producing mini speeches rather than engaging in and reacting to the unpredictable nature and flow of dynamic and emerging conversation with an interlocutor).
   - Ideally, future measurements could be constructed to include more pragmatic and discursive elements of language competencies.
   - Finally, as we have come to learn, the real value of short-term SA probably has very little to do with linguistic skills at all; SA administrators would, thus, perhaps benefit from shifting their focus to investigating the affective gains produced by SA, which are perhaps more likely to reflect the deep and far-reaching impact that SA truly does have on students' lives.

2. Concerning the Development of SA Programs for JEFLs:
   - SA administrators must clearly identify program goals from the start. Ideally, home universities and host universities would work in close conjunction to meet the needs and goals of a particular group of students.
   - To assess the efficacy of their programs (and determine whether certain objectives have been met), SA professionals need to continually monitor the development of students who study abroad (i.e., through observation and by soliciting feedback from students).
   - To maximize the SA experience for students, SA professionals should work to develop sufficient pre-SA preparation courses and post-SA follow-up sessions.
   - Pre-SA preparation courses should go well beyond initial orientation sessions (that help students obtain their passports, buy their plane tickets, process forms, etc.) to involve expectations and cultural awareness training, language training specific to the SA context, and practical and logistical information (Amelsvoort, 1999).
   - Post-SA sessions can help students reflect on their SA experience and provide them with opportunities to reinforce what they learned while abroad. Activities such as surveys and presentations would enable students to discuss (and deconstruct) aspects of their SA experience with their peers (Minehane, 2012). Furthermore, for students who have been abroad for extensive periods of time in particular, re-entry support would also be helpful (Chapell, Inaldo, White, & Pirani, 2008).

- Wherever possible (i.e., concerning both accommodation and classroom contexts), SA program organizers should try to avoid putting Japanese students in monolingual settings.
- Ideally, EFL instruction in short-term SA contexts would do well to focus on practical language usage and experiential learning rather than test preparation and lecture-based methods.
- Lastly, when dealing with Japanese EFL learners, it really helps to have SA professionals who are familiar with (and sensitive to) the general wants and needs of this particular group of learners. As professionals in this context may very well know, some Japanese EFL learners, regardless of their English proficiency level, can have an especially difficult time conveying feelings of dissatisfaction and unhappiness directly (Cutrone, 2013); thus, it is often left to the instructor and/or program director (and the receptive skills they possess) to intuitively pick up on what the students are truly feeling and to take action accordingly.

3. In terms of future research:
- To better understand the communicative needs and demands of learners in the SA context, more research comparing the L2 development of SA students with that of their AH counterparts needs to be carried out.
- In addition to controls (AH groups), which are needed to help eliminate alternative explanations of experimental results, researchers would be well served to include delayed post-tests to detect whether the effects of SA are indeed sustainable over time.
- Furthermore, researchers need to begin to assess the effects of particular teaching strategies and curriculum design on learner development in (and for) SA contexts. According to Doughty and Long (2003), Task-Based Language Teaching (TBLT) provides one such curricular framework worthy of further inquiry, as the types of interactions in which learners can engage in are particularly well-suited to SA contexts.
- Finally, concerning individual differences, little is known about how variables such as gender, age, proficiency levels, and personality affect SA outcomes and, thus, more research in such areas would be useful.

In conclusion, the present study contributes to our understanding of the SA experiences of Japanese EFL university students. In the broader context of language pedagogy, the findings of this study seem to confirm that short-term SA has little effect on students' L2 proficiency where grammar and reading are concerned. The writers hope that some of the suggestions offered in this paper will serve to assist SA professionals; however, they also fully appreciate that a great many of the recommendations provided may be beyond the control of many SA professionals, who often have to contend with various programmatic considerations as well as the practical constraints of their respective contexts. With this in mind, the writers believe that the first and most important step is to dedicate more research attention to SA and, thus, hope that this study will serve as a platform for future investigation and diagnosis into this area.

References

Amelsvoort, M. V. (1999) A pre-departure program for students who study abroad. *The Internet TESL Journal, 5*(9). Retrieved from http://iteslj.org/Lessons/Amelsvoort-PreDeparture.html

Anderson, F. (1993). The enigma of the college classroom: Nails that don't stick up. In P. Wadden (Ed.) *A handbook for teaching English at Japanese colleges and universities* (pp. 101-110). New York, NY: Oxford University Press.

Brecht, R., & Davidson, D. (1991). Language acquisition gains in study abroad: Program assessment and modification. Paper presented at the NFLC conference on language testing, Washington D.C.

Brecht, R., Davidson, D., & Ginsberg, R. (1995). Predicting and Measuring Language Gains in Study Abroad

Settings. In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 37-66). Amsterdam, Holland: John Benjamins Publishing Company.

Carroll, J. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals* I, 131-151.

Caine, N. A. (2005). EFL examination washback in Japan: Investigating the effects of oral assessment on teaching and learning. Unpublished Master's Thesis, University of Manchester, Manchester.

Chapell, L., Inaldo, J., White, M., & Pirani, C. (2008, May). Providing post-study abroad support – It's easier than you think! Paper presented at the 2008 NAFSA National Conference, Washington, DC.

Coleman, J. (1997). Residence abroad within language study. *Language Teaching*, *30*(1), 1-20.

Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition, 26*(2): 227-248.

Collentine, J. (2009). Study abroad research: Findings, implications and future directions. In M. H. Long & C. J. Catherine (Eds.), *The handbook of language teaching* (pp. 218-233). New York, NY: Wiley.

Cook, M. (2010). Offshore outsourcing teacher in-service education: The long-term effects of a four-month pedagogical program on Japanese teachers of English. *TESL Canada Journal, 28*(1), 60-76.

Cook, M. (2011, January 16). Improve teacher training at home. *The Japan Times*, p. 14.

Cutrone, P. (2009). Overcoming Japanese EFL learners' fear of speaking. *University of Reading's Language Studies Working Papers, 1*(1), 55-63.

Cutrone, P. (2013). *Assessing pragmatic competence in the Japanese EFL context.* New Castle Upon Tyne, UK: Cambridge Scholars Publishing.

DeKeyser, R. (1990). From learning to acquisition? Monitoring in the classroom and abroad. *Hispania,* 73, 238-247.

DeKeyser, R. (1991). Foreign language development during a semester abroad. In B. Freed (Ed.), *Foreign language acquisition research and the classroom* (pp. 104 – 199). Lexington, MA: D.C. Heath.

Doughty, C. J., & Long, M. H. (2003). Optimal psycholinguistic environments for distance foreign language learning. *Language Learning and Technology, 7*(3), 50-80.

Dwyer, M. M. (2004). More is better: The impact of study abroad program duration. *Frontiers: The Interdisciplinary Journal of Study Abroad, 10*(1), 151-163

Evans, M., & Fisher, L. (2005). Measuring gains in pupils' foreign language competence as a result of participation in a school exchange visit: The case of Y9 pupils at three comprehensive schools in the UK. *Language Teaching Research 9*(2), 173–192.

Freed, B. (1990). Language learning in a study abroad context: The effects of interactive and non-interactive out-of-class contact on grammatical achievement and oral proficiency. In J. Alatis (Ed.), *Georgetown University round table on languages and linguistics 1990* (pp. 459-477). Washington D.C.: Georgetown University Press.

Freed, B. (1993). Assessing the linguistic impact of study-abroad: What we currently know – what we need to learn. *Journal of Asian Pacific Communication*, *4*(1), 151-161.

Freed, B. (1995). Language learning and study-abroad. In B. Freed (Ed.), *Second language acquisition in a study-abroad context* (pp. 3-33). Amsterdam, Holland: John Benjamins.

Freed, B. (1998). An overview of issues and research in language learning in a study-abroad setting. *Frontiers: The Interdisciplinary Journal of Study-abroad*, *4*(1), 21-60.

Furmanovsky, M. (2005). Japanese students' reflections on a short-term language program. *The Language Teacher, 29*(12), 3-9.

Geis, L., & Fukushima, C. (1997). Overview of a study abroad course. *The Language Teacher, 21*(11), 15-20.

Greer, D. (2000). The eyes of Hito: A Japanese cultural monitor of behavior in the communicative classroom. *JALT Journal, 22*(1), 183-195.

Hagerman, C. (2009). English language policy and practice in Japan. *Osaka Jogakuin College Kiyo Journal*, 6, 47-64.

Hidasi, J. (2004). The impact of culture on second language acquisition. Paper presented at the Comparative

Pragmatics Association 2nd Conference, Tokyo (3 April). Retrieved from: http://www.childresearch.net/RESOURCE/RESEARCH/2006/exfile/HIDASI.pdf

Huebner, T. (1995). The Effects of Overseas Language Programs: Report on a Case Study of an Intensive Japanese Course. In B. Freed (Ed.), *Second language acquisition in a study-abroad context* (pp. 171-193). Amsterdam, Holland: John Benjamins.

Lafford, B., & Collentine, J. (2006). The effects of study abroad and classroom contexts on the acquisition of Spanish as a second language: From research to application. In M. R. Salaberry & B. Lafford (Eds.), *The art of teaching Spanish: Second language acquisition from research to praxis* (pp.103-126). Washington, D.C.: Georgetown University Press

Lee, S. I., Yoshizawa, K., & Shimabayashi, S. (2006). The content analysis of the TOEIC and its relevancy to language curricula in EFL contexts in Japan. *JLTA (Japan Language Testing Association) Journal, 9*(1), 154-173.

Llanes, À., & Muñoz, C. (2009) A short stay abroad: Does it make a difference? *System, 37*(3), 353-365.

Minehane, G. (2012). Door to door. Nagoya, Japan: Perceptia Press.

McIntyre, D. J. (2007). Preparation for long-term overseas study: Toward an integrated approach. *Asian EFL Journal, 9*(3). Retrieved from http://www.asian-efl-journal.com/Sept_2007_djm.php

Nagata, K. (2013, August 28). Forum stresses career benefits of study abroad. *The Japan Times*. Retrieved from http://www.japantimes.co.jp/news/2013/08/28/national/forum-stresses-career-benefits-of-study-abroad/#.VO-4tEIVfHg

O'Callagan, J. A. (2006). How a one semester study abroad experience affects an undergraduate college student's identity development. Unpublished Master's Thesis, Montana State University, Bozeman, Montana.

Polanyi, L. (1995). Language learning and living abroad: Stories from the field. In B. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 271–292). Amsterdam, Holland: John Benjamins.

Reesor, M. (2002). The bear and the honeycomb: A history of Japanese English language policy. *NUCB JLCC (Nagoya University of Commerce and Business Journal of Language, Culture and Communication)*, *4*(1), 41-52.

Sakui, K. (2004). Wearing two pairs of shoes: language teaching in Japan. *ELT J, 58*(2), 155-163.

Sasaki, M. (2011). Effects of various lengths of study-abroad experience on Japanese EFL students' L2 writing ability and motivation: A longitudinal study. *TESOL Quarterly, 45*(1), 81-105.

Takahashi, J. (2012). An Overview of the Issues on Incorporating the TOEIC test into the University English Curricula in Japan. *Bulletin of Tama University, School of Global Studies, 4*(1), 127-138.

Tanaka, K., & Ellis, R. (2003). Study-abroad, language proficiency, and learner beliefs about language learning. *JALT Journal, 25*(1), 63-84.

Torres, I. (2013, May 7). Government to encourage more Japanese college students to study abroad. *Japan Daily Press*. Retrieved from http://japandailypress.com/government-to-encourage-more- japanese-college-students-to-study-abroad-0728362/

Townsend, J., & Danling, F. (1998). Quiet students across cultures contexts. *English Education, 31*(1), 4-25.

Vongpumivitch, V. (2013). Assessing English in East Asia. In J. A. Kunan (Ed.), *The companion to language assessment*, Volume IV, Part 13 (pp. 1671–1680). Boston, MA: Wiley.

Yashima, T. (1999). Influence of personality, L2 proficiency, and attitudes on Japanese adolescents' inter-cultural adjustment. *JALT Journal, 21*(1), 66-86.

Yashima, T., Zenuk-Nishide, L., & Shimizu, K. (2004). The influence of attitudes and affect on willingness to communicate and second language communication. *Language Learning, 54*(1), 119-152.

Appendix A

Post-Study Abroad Questionnaire

Please answer the questions below and let us know if you have any questions.

1. Which Study Abroad Program did you attend?

2. How would you describe your overall experience abroad?

3. In retrospect, was there any information you wished you had received earlier? If so, what?

4. Please rate the academic (and on-campus) aspects of the study abroad program by circling one of the words below each sentence.

A. The preparation you received from your home university pre-departure

Poor------------Not so Good------------Satisfactory--------------Good--------------Excellent

B. The orientation and support you received from the host university upon arrival

Poor------------Not so Good------------Satisfactory--------------Good--------------Excellent

C. The accessibility and helpfulness of the host university's on-site director (and support staff) throughout the study abroad program

Poor------------Not so Good------------Satisfactory--------------Good--------------Excellent

D. The degree to which the study abroad program met your expectations overall

Not at all------------A little------------Somewhat--------------Much--------------A great deal

E. The safety of the host university's campus

Poor------------Not so Good------------Satisfactory--------------Good--------------Excellent

F. The overall classroom instruction you received at the host university

Poor------------Not so Good------------Satisfactory--------------Good--------------Excellent

G. Your instructors' course preparation and delivery at the host university

Poor------------Not so Good------------Satisfactory--------------Good--------------Excellent

H. The reasonableness of assignments at the host university

Poor------------Not so Good------------Satisfactory--------------Good--------------Excellent

I. The usefulness of assignments at the host university

Poor------------Not so Good------------Satisfactory--------------Good--------------Excellent

J. The degree to which classroom instruction at your host university matched your English proficiency level

Not at all------------A little------------Somewhat--------------Much--------------A great deal

K. The degree to which you think your English communication skills improved overall

Not at all------------A little------------Somewhat--------------Much--------------A great deal

L. The degree to which you think your TOEFL score improved

Not at all------------A little------------Somewhat--------------Much--------------A great deal

5. Please rate the following aspects of your experience staying with a host family (Note that since members of Group D did not stay with a host family, the term "host family" was substituted with "student dormitory' in this section of their questionnaire).

A. The degree to which the homestay experience met your expectations

Not at all------------A little------------Somewhat--------------Much--------------A great deal

B. The cleanliness of the home you stayed in

Poor------------Not so Good------------Satisfactory--------------Good--------------Excellent

C. The friendliness of your host family

Poor------------Not so Good------------Satisfactory--------------Good--------------Excellent

D. The food your host family prepared for you

Poor------------Not so Good------------Satisfactory--------------Good--------------Excellent

E. The safety of the home and area that you stayed in

Poor------------Not so Good------------Satisfactory--------------Good--------------Excellent

F. The support you received from your host family regarding your commute to and from your host university

Poor------------Not so Good------------Satisfactory--------------Good--------------Excellent

G. The degree to which your host family was interested in Japan and Japanese culture

Not at all------------A little------------Somewhat-------------Much-------------A great deal

H. The degree to which you were able to communicate in English with your host family

Not at all------------A little------------Somewhat-------------Much-------------A great deal

I. The degree of confidence that you will keep in touch with your host family in the future

Not at all------------A little------------Somewhat-------------Much-------------A great deal

J. The degree of support you received from your host university regarding any issues you had with your accommodations

Not at all------------A little------------Somewhat-------------Much-------------A great deal

6. Please rate the following aspects of your cultural experience while studying abroad:

A. The degree to which the overall cultural experience met your expectations

Not at all------------A little------------Somewhat-------------Much-------------A great deal

B. The sociocultural activities you participated in, as organized by your host university

Poor------------Not so Good------------Satisfactory-------------Good-------------Excellent

C. The sociocultural activities you participated in, as organized independently and informally by you and/or your friends

Poor------------Not so Good------------Satisfactory-------------Good-------------Excellent

D. The sociocultural activities you participated in, as organized by your host family

Poor------------Not so Good------------Satisfactory-------------Good-------------Excellent

E. The degree to which you were able to learn about and experience the host culture

Not at all------------A little------------Somewhat-------------Much-------------A great deal

F. The amount of contact you had with people in the local community off-campus (not including host family)

Not at all------------A little------------Somewhat-------------Much-------------A great deal

G. The degree to which you were able to communicate successfully in English with people in the local community off-campus (not including host family)

Not at all------------A little------------Somewhat-------------Much-------------A great deal

H. The degree to which you were able to make friends with people in the local community off campus (not including host family)

Not at all------------A little------------Somewhat-------------Much-------------A great deal

7. What aspects of your host university's study abroad experience did you enjoy the most? Please explain your answer.

8. What aspects of the study abroad experience did you dislike and/or struggle with? Please explain your answer.

9. Would you recommend the Study Abroad Program you participated in to other students? Why or why not?

10. Generally speaking: If you could change anything about the study abroad options offered in your home university, what would it be? Please explain your answer.

Appendix B

Questionnaire for AH Students

Please answer the questions below and let us know if you have any questions.

1. Why did you not study abroad this summer?

2. How much did you study and/or use English this summer?

Not at all------------A little------------Somewhat-------------Much-------------A great deal

3. Relating to your answer to Question 2 above, please describe specifically what types of activities you did this summer to improve your English?

4. How much did you study specifically for the TOEFL test this summer?

Not at all------------A little------------Somewhat-------------Much-------------A great deal

5. Relating to your answer to Question 4 above, please describe specifically what types of activities you did this summer to improve your TOEFL score?

Practice of Differentiated Instruction and Alternative Assessments with Young EFL Learners in Taiwan

Yu-ju Hung*
R.O.C. Air Force Academy, Taiwan

Abstract

Drawing on Vygotsky's (1978) sociocultural theory, the present study took a classroom-based research approach to investigate the practice of differentiated instruction (DI) with 26 second-graders in an English-as-a-Foreign-Language (EFL) class in southern Taiwan. The teacher-researcher scaffolded students' learning by modifying tasks in response to various ability levels. The teacher introduced the major content in the textbook supplemented with a picture book. A variety of multi-leveled activities were then conducted, including a heterogeneous grouping role-play, a homogeneous grouping out loud sentence reading, phonics practice activities, along with other alternative assessments. The lesson was finalized by providing tiered worksheets for students to self-select and complete an assignment. In congruence with the sociocultural tenet that learners perform higher function thinking through interaction with more knowledgeable others, the results show that the majority of the students engaged in the collaborative work, eventually challenged higher level tasks, and completed the worksheets correctly. In the survey administered after DI, the students generally perceived the learning experiences positively, with 80% satisfied and only 5% dissatisfied. However, the number of lower level students who showed satisfaction was significantly lower than the number of higher level students. The instructor also observed that lower level students tended to be off-task if they became frustrated with course content, suggesting the need for extra facilitation or remedial teaching for lower level students. Implications for future practice and research are discussed.

Keywords: differentiated instruction, young learner, alternative assessment, sociocultural theory

Introduction

English remains the major language of international communications, the growing importance of which necessitates its use among non-native speakers of English across national and cultural boundaries. A trend in East Asia is to move English education to a younger age for students. In Taiwan, to develop students' English abilities, the English curriculum that was originally introduced in the fifth grade was shifted to the third grade in 2005 (Chen & Tsai, 2012; Chern, 2002). However, some parents choose to start their children's English education in bilingual kindergartens and continue to have them learn English in private English institutes. While some students enjoy this privilege, others are beginners upon entering elementary school (Liu, 2008). Under these circumstances, one-size-fits-all classrooms often end up dis-servicing students: middle-level students show average achievement, lower-level students suffer, and higher-level students make few gains (Schumm, Moody, & Vaughn, 2000). The big gap between students' proficiency levels, the "two-peak phenomenon" to use Chien's (2012) phrase, makes English teaching more difficult than ever and highlights an urgent need for effective teaching to accommodate a diverse range of students.

* *Tel. 886 7 6268827; Email: hung.yuju@gmail.com; Foreign Languages Section, General Education Center, R.O.C. Air Force Academy, Taiwan*

To respond to all learners' needs, the practice of differentiated instruction (DI) has been introduced as a way of "ensuring that what a student learns, how he/she learns it, and how the students demonstrate what he/she has learned is a match for that students' readiness level, interests, and preferred mode of learning" (Tomlinson, 2004, p. 188). Unlike traditional instruction in which teachers "teach to the middle" (Haager & Klingner, 2005, p. 19), DI addresses not only the current level of each learner but also individual needs, interests, and strengths so as to move them forward as far as possible (Gregory & Chapman, 2002; Heacox, 2002; Levy, 2008; Tomlinson, 2005). DI has been used in both literacy and content courses in first language (L1) settings (Dahlman, Hoffman, & Brauhn, 2008) whereas the application of DI in second language (L2) settings has been limited. Therefore, the purpose of this study was to investigate the effects of DI along with group and peer assessments in a target English-as-a-Foreign-Language (EFL) course in an elementary school in Taiwan and explore how the instructor and students perceived this approach to teaching in this context.

Conceptual Framework

Vygotsky's (1978) social constructivist theory, which puts a central focus on social interaction, serves as theoretical support for differentiated instruction (Subban, 2006). Vygotsky believed learning occurs on two levels: "First, on the social level, and later, on the individual level; first, between people (interpsychological) and then inside the child (intrapsychological).…All the higher functions originate as actual relationships between individuals" (Vygotsky, 1978, p. 57). In DI, teacher-student interaction takes place as the instructor scaffolds and models the tasks while student-student interaction increases in heterogeneous as well as homogeneous grouping activities.

Vygotskian theory proposes social processes as a mechanism for learning and suggests that learners' thinking and meaning-making emerge out of their social interactions with those around them in the "zone of proximal development" (ZPD) (Kaufman, 2004). The ZPD is defined as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (Vygotsky, 1978, p. 86). The ZPD supports higher-level cognitive development than the learner is able to achieve alone, and productive interactions orient learning toward the ZPD (Palinscar, 2005). In a DI classroom, the instructor plays a guiding role by diagnosing students' differences and then planning lessons and teaching methods accordingly. Teachers in DI classrooms guide students by being flexible in adjusting the content, process, or product of teaching and learning based on pre-assessments (Lewis & Batts, 2005; Nordlund, 2003; Tomlinson, 1999, 2005). Regarding content, teachers provide different content to meet the needs of individual students. For process, a variety of multi-level activities and strategies are used based on students' interests, learning styles, and cognitive capacities. The product is differentiated by offering choices of ways for students to demonstrate their learning outcomes (Heacox, 2002; Tomlinson, 1999, 2005). With that being said, DI "requires extensive knowledge as well as a vast repertoire of instructional strategies ready to match each and every student's different needs" (King-Shaver & Hunger, 2003, p. 2).

Wertsch (1991) identified semiotic mediation as the second theme of Vygotskian theory. Human action is mediated by tools and signs, including "language; various systems of counting; mnemonic techniques; algebraic symbol systems; works of art; writing; schemes, diagrams, maps and mechanical drawings; all sorts of conventional signs and so on" (Vygotsky, 1981, p. 137). Halliday (1993) proposed that "the prototypical form of human semiotic is language" (p. 93), and Wells (2007) argued that dialogue is the most powerful among artifacts mediating activities which facilitate the co-construction of knowledge. In addition to providing choice and constant adjusting of content, process, and products to meet all students' needs, the principles of DI include a focus on the integration of authentic assessment and instruction and flexible grouping (Rock, Gregg, Ellis, & Gable, 2008; Tieso, 2003; Tomlinson, 1999). Students in a DI classroom engage in various forms of group activities. By interacting and discussing with their peers in meaningful tasks as well as participating in group and peer assessments, students are able to conduct high levels of cognitive thinking, build knowledge together, and eventually develop English ability.

Grounded on sociocultural theory, this study was designed to investigate how students and their teacher perceived the tiered tasks scaffolded and facilitated by the teacher. Primary data included the results of a student survey, the number of tasks completed, the teacher's teaching notes, and a teacher interview. In addition, observations of students' collaborative work in a role-play activity, picture book reading, phonics learning, and their responses in a peer assessment were analyzed.

Differentiated Instruction

Substantial evidence now shows the benefits of Differentiated Instruction as an effective teaching philosophy for improving student learning in heterogeneous L1 classrooms in various fields such as math (Tieso, 2005), science (Mastropieri et al., 2006; Odgers, Symons, & Mitchell, 2000), language arts (Ankrum & Bean, 2008; Coulter & Groenke, 2008), and special education (Lawrence-Brown, 2004). While DI has started to draw the attention of educators and researchers in EFL settings, little work in this area has been reported so far.

The main focus of DI research in EFL settings can be divided twofold, on different learning styles or on mixed achievement levels. In regards to students' learning styles, Alavinia and Farhady (2012) investigated whether DI has any impact on Iranian EFL learners' vocabulary learning. Sixty students who were homogeneous according to a pre-assessment were divided into an experimental class and a control class. The students in the experimental group were formed into groups according to the results of a multiple intelligences test and a learning style test. Students who had the same types of intelligences and learning styles received differentiated instruction matching their type. Comparison of English pre-test and post-test scores of students in both classes showed DI had some positive impact on the students' learning. Also, students in the DI class felt more comfortable and confident with their performance. However, the results within each group remained questionable since there was no significant difference between the performance of learners with different learning styles, but there was a significant difference between the performance of learners with different intelligences.

Contrary to the results yielded in Alavinia and Farhady's (2012) study, the findings in Alavinia and Sadeghi's (2013) study showed that DI responding to learners' styles did not enhance students' learning outcome. Forty-seven EFL freshmen at Urmia University who had similar TOEFL scores participated in this study. Twenty-four of them were randomly assigned to the experimental group, and 23 were assigned to the control group. All of the students took a learning styles test. Each group consisted of an equal number of each learning style and gender type. DI was implemented in the experimental group whereas task-based instruction was given in the control group. The same TOEFL test was administered and no significant differences were found between the two groups.

Other than responding to learners' learning styles, DI has been applied in mixed-level EFL classrooms to match students' language levels to instruction. Chen (2007) examined how DI was applied in college EFL classes in Taiwan and how students perceived tiered performance tasks. Twelve freshmen in the Department of Applied Foreign Language participated and were offered three-leveled tasks, weighted according to level of difficulty, from which to choose as part of their final examination grades. The interview data showed that the students enjoyed taking ownership to decide among the leveled tasks and agreed that this empowering process would improve their English skills. However, some students were concerned that higher-leveled tasks increased their anxiety. To take advantage of providing students task choices and at the same time to avoid causing anxiety, the teacher carefully designed tiered tasks suitable for students' current levels. Students were assured that everyone would be able to learn at their own pace, choose tasks they were comfortable with, and challenge themselves with another level if they wanted to.

Zhan and Chen (2012) investigated whether DI has a positive impact on students' English word recognition, reading comprehension, and reading attitudes. Fifty-eight third graders from northern Taiwan, 29 of whom were in the experimental group and received balanced reading instruction, which integrated skill-oriented and meaning-based approaches. Based on their pre-assessed English levels, the students were grouped homogeneously. The class started with an introduction of reading to the whole group. Then, the teacher gave separate instruction to each group, generally spending more time with lower-level students since they needed more practice. Then the whole class came together to share and demonstrate what they had learned. The control group followed the traditional approach of teaching the same content to the whole class, pointing out new vocabulary, and choosing several students to check their pronunciation. Then questions were asked to confirm students' understanding of the story. Finally, the students read the story to each other and then some of them read in front of the class. The results revealed that the experimental group did not perform better than the control group in terms of word recognition, but outperformed them in reading comprehension and reading attitudes. To build upon Zhan and Chen's (2012) course design, this study increased complexity by including not only reading, but also speaking and phonics exercises. Different grouping forms were also implemented as reference for future practice.

Chien (2012) implemented DI with a group of sixth graders in an EFL class. Instruction was differentiated in terms of content, process, and product. For example, six riddles were selected as supplementary materials to the textbook and allowed lower level students to guess the words in Chinese while higher level students usually responded in English. On a test of word recognition, lower level students only had to recognize easier words while higher level students were required to recognize more difficult words and to spell them correctly. Several other tasks based on students' various English levels were offered for the students to choose. Another example of DI that was implemented was giving extra instruction to lower level students. Chien concluded that DI made English instruction more accessible to students. Students could choose their own tasks and learn at their own pace. The author also pointed out that to be successful, DI required not only careful lesson design and delivery but also a well organized class routine.

Among the studies showing an initial success of DI implementation in the EFL context in Taiwan, both targeted young learners (Chien, 2012; Zhan & Chen, 2012) and emphasized giving extra instruction to lower level students. Their research confirms that students at different levels need different kinds of attention and methods of instruction, highlighting the importance of student perspectives. Since meeting the needs of each individual student is the main goal of DI, it is essential to explore how students at various levels perceive the activities and assessments. Therefore, the present study expands this focus to investigate and compare the perceptions of higher and lower achievers. Also, to bridge the gap that most studies have reported with experimental research, this study takes a classroom-based research approach to move from the outcomes of DI to the specifics of how it operates in a real situation and the challenges it presents to teacher and students.

Method

Research Design

A classroom-based research approach was employed in this study. An elementary EFL teacher, who directly experienced the benefits and challenges of DI worked collaboratively with the author, a university researcher, on the planning and implementation of DI procedures. At the end of the DI period, instructor interview questions were developed and conducted by the author. This paper reports this primary stage of the DI implementation.

Participants and Context

The setting of this study was a public elementary school in southern Taiwan. The school was established in 1996 to serve a new high socioeconomic status (SES) suburban community and was regarded as a high performing school. Teachers were encouraged to use non-traditional teaching approaches to increase students' learning motivation and outcomes. The instructor learned about DI in a professional development training session, was fascinated by the approach, and decided to try it in her class.

The total student population of this school was about 800 students, divided into 30 classes (grades 1-6). Approximately 90% of the students were Taiwanese; 10% were Hakka (an ethnic Chinese group comprising 15-20% of Taiwan's population) or immigrants. When the study was conducted, the national policy required students to learn English from the fifth grade. However, the local educational policy promoting English proficiency required all students to start English courses in the second grade.

Twenty-six second graders taught by Author 2 joined the research, 11 female students and 15 males. Approximately 70% of them (N=18) were also learning English at private language institutes after school, and 53% (N=14) had started to learn English in kindergarten. Therefore, some of the students had already learned the main textbook content whereas a few still had problems with letter names and letter sounds.

Procedures for DI

This study took place over one semester in which the students received two 40-minute English classes every week. Pre-assessment and wrap-up evaluations each lasted one week. The role-play activity took four weeks. Picture book reading and phonics exercises each took two weeks. For the rest of the classes, the teacher focused on the textbook content.

Pre-assessments

The teacher wanted to learn how much lower achievers had lagged behind in the previous semester and to what extent the higher achievers already knew the current content. In the pre-assessments that the teacher designed, a poster that illustrated the context of the dialogue in Unit 1 was presented to the students. Each student in turn said the English words that the poster suggested, which the teacher wrote down. Based on the average scores of their pre-assessment and their final grades in the previous semester, the students were divided into three levels: high (eight students), mid (nine students), and low (nine students). The students were aware of their own levels. Concerned that students assigned to the low level might feel a loss of face, the teacher assured them it was normal for students to learn at different paces, and they all had their own strengths and weaknesses in different subjects. According to our observations, the students all felt fine with their levels.

DI Activity I: role-play

The class started with whole class instruction of the textbook, *Enjoy 2*. The target sentence structures were, "Who's he/she? He's/She's ______ (name or family member)." After introducing the dialogue, which included these sentence structures, the teacher flashed word cards and the whole class did drills. Then a song that contained these sentences was taught. The students practiced the song as a whole class, in groups, and individually; then, the students read every word on the word cards by themselves and self-evaluated how many words they could say.

Table 1
Dialogue Sample Adapted from the Textbook

Lesson One: Who's He?
Ingrid: Look! That's Quincy. Mrs. Wu: Who's he? Ingrid: He's my friend. Mom, this is Quincy. Mrs. Wu: Nice to meet you, Quincy. Ingrid: Quincy, this is my mom. Quincy: Nice to meet you, too, Mrs. Wu.

adapted from Enjoy 2

Students were then grouped with each group being made up of three students of different levels. The instructor explained to the students the goal of the heterogeneous grouping. The group members should learn collaboratively. They could create their own dialogue based on the sample dialogue in the textbook (Table 1). When doing a role-play, a higher level student could take charge of more complex sentences, and a lower level student could be responsible for easier sentences. The teacher facilitated each group when needed. When ready, each group performed a role-play in front of the whole class; each presentation was evaluated by the teacher and the students' peers (Table 2). Though students in the same group were in charge of sentences of various complexity, the teacher's encouragement of collaboration enabled them to discuss their dialogue together. The high level students automatically played the role of leaders to lead the discussion and help their group members. Regarding the peer assessment activity: though the students appeared to be excited about evaluating their peers, their written comments were limited which was probably due to their young age. Some of the groups failed to write down the numbers of the groups they assessed, so grades could not be calculated. These results demonstrated that the procedure of peer assessment needs to be simplified for young learners.

Table 2
Peer Group Evaluation Form of the role-play Activity

criteria	ratings		
content	☺	☺	☺
voice, fluency	☺	☺	☺
facial expression, body language	☺	☺	☺
collaboration	☺	☺	☺
props, costume	☺	☺	☺
Total			
Strength			
Suggestion			

DI Activity II: Picture Book Reading

The picture book *I Love My Family* was selected as supplementary material because the book content was similar to that of the textbook. It supported multi-leveled extensive learning. Lower level students used it as a review of words of family members (Task A). Mid level students learned the extensive sentences, "I love my ____. My _____ loves me." (Task B). In addition to these, higher level students were expected to learn the last sentence, "He _____ (flies kites) with me." (Task C). The reading aloud activity was incorporated because reading a text aloud has been found to benefit learners' word recognition (Magno & Sison, 2011). After the teacher introduced the picture book, students chose the level of the task they wanted to practice. Students formed groups with the students who had chosen the same level of the tasks. Each group practiced together, while the teacher went to each group and provided help; then, each group did a within group peer evaluation. Each group member took turns reading the target words or sentences on each page of the picture book. The other group members listened and marked how many pages the students read correctly. Finally, the teacher collected the evaluation forms.

DI Activity III: Phonics

The teacher used the website Starfall, an educational website providing phonics and basic reading materials for young learners, to review letter names, letter sounds, and example words. Then, initial sounds of the example words were repeated and stressed to introduce concepts of vowels and consonants, such as *g*, /g/, *girl*. The students practiced as a whole class using the sounds and animations on the website. An alphabet poster was displayed to provide another opportunity for students to practice again. After the whole class instruction, students chose the task that fit their level and formed pairs with another student selecting the same task. In Task A, students read each letter's name and its phonics out loud, such as *g* /g/. Task B required students to read the word for the picture that illustrated each letter out loud. In Task C, students read a sentence using the word matching each letter out loud. After the practice, each student performed the selected task again, and the partner listened and marked on the form that had the same content as the alphabet poster.

Wrap-up Evaluations

Two types of tiered tasks were designed from which students could select those that matched or challenged their present level. For vocabulary and spelling, Task A asked students to match pictures with family members. Task B asked students to write the first letter of the word for each family member. Task C required students to spell the words. Task D required students to spell the words and write another word related to the family member.

For recognizing and writing the initial sounds, again four tasks were provided. In Task A, students heard a sentence and wrote down the initial sounds of emphatic words in the sentence. They could open the textbook to refer to the content. For example, after reading "Grandpa, grandma, let's go," the teacher repeated and emphasized "grandpa, grandma, go." In Task B, students did the same thing, but could not refer to the textbook. In Task C, students not only wrote the initial sounds but also spelled the words. In Task D, students wrote the initial sounds and the whole sentences. For these two final evaluations, students could accept the challenge of an upper level task if they finished the task at their current level. The teacher intentionally tried providing tasks at four levels instead of three. The students were found to appreciate the choice and enjoy challenging themselves at different levels.

From the teacher's observation and students' feedback, the DI process was found to be rewarding. However, the teacher also found it was difficult to select materials that were challenging but still approachable to students as well as materials that met low level students' abilities but were not too easy for them. Second, she needed a large repertoire of teaching strategies to design tiered tasks and worksheets. Third, when she proceeded to another task or did different grouping, it was rather difficult to guide her students through the transition.

Data Sources and Analyses

Student Survey

A three-level Likert scale questionnaire was designed to assess students' perceptions of the DI experience. The first part elicited students' background information, including their gender, whether they learned English outside school, and when they started to learn English. Items 1 to 3 explored how they liked the role-play activity, and Items 4 to 6 examined how this activity benefited their English learning. Items 7 to 10 elicited students' attitudes toward task choice in the activities of picture book reading and phonics practice. Items 11 to 13 were concerned with how these opportunities to select their own tasks influenced their learning (Table 3). The survey was administered in Chinese. Because the participants were young learners and to avoid influencing their responses, the teacher stood in front of the class, provided an explanation and guidance for each item, and assured the students that their answers would not influence their grades in any way.

Table 3
Survey of Students' Perception about DI (English translation)

When I do the role-play activity,
1. I like the role-play group activity.
2. I am able to collaborate with my classmates.
3. I am glad to work with classmates with different levels of English abilities.
4. I can learn actively.
5. I can improve my English.
6. I am more confident in speaking English.

When I choose my own tasks in the activities of picture book reading and phonics practice,
7. I like to decide my own tasks.
8. I am able to choose the tasks appropriate for me.
9. I like English class more.
10. I am able to collaborate with my classmates.
11. I can learn actively.
12. I can improve my English.
13. I am more confident in speaking English.

Teaching Notes and Teacher Interview

The teacher kept a record of their teaching reflections. To elicit the teacher's opinions more extensively, an interview was conducted to determine their perceptions of DI. Items 1 to 3 of the interview protocol captured the teacher's notion of DI. Items 4 and 5 elicited their experience of practicing DI. Items 6 to 8 were about their reflections and ideas for further implementation of the approach (Table 4). The interview was administered in Chinese and lasted about one and a half hours. The whole process was recorded and then transcribed and translated by the author.

Table 4
Teacher Interview regarding DI Practice (English translation)

1. How did you find out about DI? What does DI mean to you?
2. Why did you want to practice DI in your class?
3. How did you explain DI to your students?
4. How did you perceive this practice?
5. Did you encounter any challenges when you practiced DI?
6. Based on your observation, how did your students like DI? Did high level students react differently from low level students?
7. If you practice DI again, what will you do differently?
8. Do you have any suggestions for other English teachers who want to practice DI?

Peer Assessment Survey

Table 5 shows the questions designed to uncover students' perceptions of peer assessment of the role-play activity. The questionnaire was a three-level Likert scale, in which students chose whether they were happy, neutral, or unhappy about each statement. Items 1 to 4 asked whether students liked to assess their peers and being assessed. Items 5 to 8 concerned students' ability to conduct peer assessment. How peer assessment helped with English learning was explored in Items 9 to 12. Items 13 to 16 concerned group collaboration. The administration procedure was the same as the above student survey.

Table 5

Peer Assessment Survey (English translation)

When I do group peer assessment,
1. I like to grade my classmates.
2. I like my classmates grade me.
3. I like to give my classmates comments.
4. I like my classmates to give me comments.
5. I am able to grade my classmates.
6. My classmates are able to grade me.
7. I am able to give my classmates comments.
8. My classmates are able to give me comments.
9. When I grade my classmates, I will pay attention to their performance.
10. Watching my classmates' performance is helpful to my English learning.
11. Grading each other is helpful to my English learning.
12. My classmates' comments are helpful to my English learning.
13. Group discussion of grading is helpful to my English learning.
14. When discussing grading in groups, I like to discuss with my group members.
15. When discussing grading in groups, I have opportunities to say my opinions.
16. When discussing grading in groups, I will listen to my group members' opinions.

All of the worksheets and evaluation forms completed by the students were collected as supporting data. Quantitative data were analyzed using descriptive statistics and presented along with themes that emerged from the qualitative data.

Results

Due to the small number of participants, only percentages and numbers are reported here instead of more advanced statistical analyses. Rather than a conclusion, the results represent initial findings to be used as a reference for future teaching practices and related research.

Students' Perceptions of DI

Table 6 shows the number of students who selected each tiered task. Before the study, we wondered whether students might choose easier tasks than their levels warranted, but the distribution across tiers dispelled this concern. For each activity, only half the students chose the lower two tasks, and the least number of students chose the easiest task. Across three activities, the number of students who chose Task C or Task D was the same (N=13) with the majority choosing the most challenging task available. The numbers of the students who chose Task A or Task B varied in each activity, which might imply that low and mid level students were aware of their competence to accomplish the task in each activity and tried to be cautious in selecting a task. Some of the low level students started from the easiest task and moved on to a higher level. Also, these students were confident about recognizing initial sounds, so only one student selected the easiest task in this activity.

As displayed in Tables 7 and 8, the students generally perceived this DI experience positively. Table 7 shows that 73% of the students were generally satisfied with the role-play activity. The students were able to collaboratively learn (Items 2-3), and this activity helped to enhance their English ability (Items 4-6). Compared to the rest of the items in Table 7, fewer students were satisfied (69%) with heterogeneous group work, with an increase in those that were unsatisfied (12%). As this activity seemed to be the least well received, further investigation of students' responses to it should be made.

Table 6
The Numbers of Tasks Completed

picture book reading		vocabulary learning		phonics practice	
Task	Students	Task	Students	Task	Students
A	3	A	5	A	1
B	10	B	8	B	12
C	13	C	4	C	4
		D	9	D	9

Note. N = 26; Task A is the easiest task, and Task D is the hardest task.

Table 7
Students' Perceptions of role-play

	happy		neutral		unhappy	
	N	%	N	%	N	%
• general satisfaction	19	73%	7	27%	0	0%
• ability to collaborate	20	77%	4	15%	2	8%
• heterogeneous grouping	18	69%	5	19%	3	12%
• active learning	20	77%	6	23%	0	0%
• English improvement	20	77%	5	19%	1	4%
• confidence in speaking English	21	81%	4	15%	1	4%

Note. N = 26

As shown in Table 8, 77% of the students were generally satisfied with being offered choices of tasks in the activities of picture book reading and phonics practice. The majority of students enjoyed the activities and could work in a homogeneous group. They also indicated these activities helped with their English learning (Items 11-13). It is noteworthy that the highest number of positive responses (96%) was for Item 8, the ability to choose tasks. This positive response shows students appreciated being empowered to select their own tasks.

Table 8
Students' Perceptions of Choice of Tasks

	happy		neutral		unhappy	
	N	%	N	%	N	%
• general satisfaction	20	77%	4	15%	2	8%
• ability to choose tasks	25	96%	0	0%	1	4%
• like English class	20	77%	5	19%	1	4%
• ability to collaborate	22	85%	3	11%	1	4%
• active learning	18	69%	7	27%	1	4%
• English improvement	23	88%	1	4%	2	8%
• confidence in speaking English	23	88%	3	12%	0	0%

Note. N = 26

Table 9
Perceptions of Students in Different Levels

	high (N=8)	mid (N=9)	low (N=9)
role-play			
1. general satisfaction	2.88	2.78	2.56
2. ability to collaborate	2.88	2.78	2.44
3. heterogeneous grouping	2.88	2.33	2.67
4. active learning	2.75	2.89	2.67
5. English improvement	2.88	2.89	2.44
6. confidence in speaking English	2.63	2.78	2.89
choice of tasks			
7. general satisfaction (task choice)	2.88	2.44	2.78
8. ability to choose tasks	3.00	2.89	3.00
9. like English class	3.00	2.56	2.67
10. ability to collaborate	2.75	2.89	2.78
11. active learning	2.88	2.78	2.33
12. English improvement	2.75	2.89	2.89
13. confidence in speaking English	3.00	2.67	3.00
Average	2.86	2.74	2.70

Note. N=26. "Happy" was coded as 3, "neutral" as 2 and "unhappy" as 1.

Table 9 compares the responses of students with different proficiency levels. The satisfaction level of students in the low level group was slightly lower than that of the students in the mid and high level groups. The satisfaction level for the high level group was above 2.6 across all the items. Mid level students were the least satisfied with heterogeneous grouping (Item 3=2.33) and task choices (Item 7=2.44). Low level students had relatively more problems with collaboration (Item 2=2.44), English improvement (Item 5=2.44), and active learning (Item 11=2.33).

Table 10 explores students' experience of peer assessment in the role-play activity. Interestingly, while 85% of the students reported being happy with the activity (Item 1) and 84% considered themselves capable of assessing each other (Item 5), only 69% reported they liked giving comments (Item 3) or felt able to do so (Item 7). Also, while 89% liked being given comments (Item 4), and 85% thought others were able to grade them (Item 6), only 65% thought that their peers' comments were helpful to their English learning (Item 12). Such apparent contradictions warrant further study. It should also be noted that while 81% of the students enjoyed group discussion (Item 14), only 73% found it useful to their English learning (Item 13), only 69% were happy about voicing their opinions (Item 15) and only 73% about hearing the opinions of others (Item 16), another apparent discrepancy that should be explored. Overall, the somewhat ambivalent findings suggest that these young students were still getting used to discussion and collaboration as a way of learning and needed strong teacher support.

Table 10
Peer Assessment Survey

	happy		neutral		unhappy	
	N	%	N	%	N	%
1. like grading others	22	85%	3	11%	1	4%
2. like being graded	21	81%	3	11%	2	8%
3. like giving comments	18	69%	7	27%	1	4%
4. like being given comments	23	89%	3	11%	0	0%
5. able to grade others	22	84%	2	8%	2	8%
6. others able to grade me	22	85%	4	15%	0	0%
7. able to give comments	18	69%	6	23%	2	8%
8. others able to give me comments	25	96%	1	4%	0	0%
9. paying attention to others' performances	21	81%	4	15%	1	4%
10. watching performance helpful to English learning	20	77%	5	19%	1	4%
11. grading helpful to English learning	20	77%	5	19%	1	4%
12. peers' comments helpful to English learning	17	65%	7	27%	2	8%
13. group discussion helpful to English learning	19	73%	6	23%	1	4%
14. like to discuss in groups	21	81%	4	15%	1	4%
15. saying my opinions	18	69%	5	19%	3	12%
16. listening to others' opinions	19	73%	6	23%	1	4%

Note. N=26

Teacher's Perception of DI

Commenting on the positive impact DI had on her students, the teacher distinguished between traditional teaching, which she felt did not help either high- or low-level students, and DI, which the teacher found to be a way to provide every student with achievable English tasks and, therefore, successful experiences. As she stated below:

> After practicing DI, I realize higher achievers can do more, and lower achievers can complete the tasks at their own levels. Everyone has pleasant learning experiences in this way. That's why I want to continue to practice DI. When students get a sense of achievement from learning, they keep on learning.

According to the teacher, providing a choice of tasks to challenge students and convincing students to accept DI are the keys of a successful DI classroom. After different worksheets were shown to students, each student selected one that met their current level. As the teacher described, "students showed pride on their faces. Several students chose to challenge themselves at a higher level after they were done with the first worksheet." When asked how DI was introduced to their students, the teacher used running as a metaphor: "It's like running. Some run faster, and some run slower. It doesn't matter where you are now. Only if you keep running will you get to your own destination."

In the teaching notes, the teacher indicated that DI embodied the essence of learner-centered teaching, in which the word "learner" means "every learner." When she planned their lessons, she took the needs of each student into account. The course content as well as the assessments should be multi-leveled. The teacher carefully ensured that the most difficult tasks were still achievable by top level students to avoid frustration. She observed that DI improved students' learning attitudes; when she saw that students engaged in activities and group work, she realized that a teacher's role was now to facilitate, and no longer at the center of all learning.

Because every student was accommodated and learning attentively, the teacher did not need to do as much remedial teaching as before. Before DI was implemented, the teacher found that some students always lagged behind. Usually the teacher provided extra instruction after class to help these students keep up with others. In a DI classroom, every student was learning. As long as the teacher paid attention to the learning of low level students and offered timely facilitation, extra remedial teaching was not as urgently needed.

The teacher also found that traditional assessment did not fit in with DI. The purpose of assessment was not just to evaluate students, but to provide a learning opportunity for students as well as to help the teacher understand where the students were in their learning. The teacher could constantly adjust their instruction based on students' performances in the assessment. In addition, multi-task assessment offered a channel for students to put forth more effort and challenge themselves. The teacher was excited to see that happen:

> In the past, assessment was based on students' academic achievement. In a DI classroom, we evaluate students' effort. If students moved from a lower level task to a higher level task, they deserved credits to reward their effort.

Though the experience was generally positive, the teacher also found that lower achievers easily went off-task when grouping was homogeneous. This phenomenon did not occur in high and mid level groups. The students in low-level groups tended to be lower motivated learners and gave up easily when facing difficulties. The teacher considered it essential to provide immediate help for them, or avoid grouping low achievers together. On the other hand, low achievers participated in tasks actively when grouping was heterogeneous. A higher level student usually played the role of a group leader to guide and help the other group members.

Moreover, the instructor would be more careful with grouping in their future practice. The students were assigned to groups for role-play based on the results of pre-assessments. In one of the groups, the students could not get along with each other and argued. After the teacher changed one of the group members with a student from another group, the collaborative work went smoothly. Though the original plan was grouping based on English levels, the teacher would take into account students' personalities and social relationships. Group members were moved around or students were allowed to form their own groups as long as they found three students at different levels for heterogeneous grouping. Also, explicit instruction about collaborating with peers should be given.

Discussion and Conclusion

In accordance with the positive results reported in related studies (Chen, 2007; Chien, 2012; Zhan & Chen, 2012), the majority of the students as well as the teacher also reported positively on this DI experience, which involved collaboration between teacher and students as well as among students, thus supporting Vygotsky's (1978) sociocultural theory that learning takes place first between people, and then within the learners. In this study, the teacher first played a guiding role, leading the learners to do collaborative work in groups, followed by individual task choice activities. She carefully introduced this student-centered approach to the class while guiding students away from the teacher centeredness that is still the norm in many mainstream classrooms, informing them of the purpose and goal for each activity. The young learners were assured it was acceptable for each to learn at their own pace, so resistance to this new learning experience was minimized from the beginning.

The second key to success was that appropriate materials and tiered tasks were carefully selected or designed by the teacher, who used their expertise to guide students step by step to avoid chaos when switching to new activities, though it still took some time to get students on-task. Also, instead of traditional written tests, various types of alternative assessments played an essential role at different stages in DI. The results of the assessments informed the teacher of all students' learning progress, so she could modify their teaching accordingly. As seen in progressivism, which Sivasubramaniam (2011) considered a learner-centered educational ideology as well as a lead-in to constructivism, a personalizing approach to English learning should be prioritized. Instead of homogenizing students in norm-referenced assessment, a personalizing approach encourages teachers and students to become involved in and negotiate the assessment process. By taking a participatory role, learners are empowered and gain autonomy in the learning process whereas teachers are informed and become reflective practitioners.

Though the teacher was satisfied with this first trial, she pointed out that looking for suitable resources and designing multi-level activities took much more time than before. Professional training in the principles of DI and activity design would be useful. Also, building a library of English resources offering various topics and difficulty levels as well as textbooks that came with tiered worksheets, extensive supplementary materials, and online resources would expedite the implementation of DI for teachers.

In addition to the teacher's scaffolding, the students learned from interacting with peers in various groupings. The teacher was gratified to see that most students were engaged in group work and learned from each other. However, the teacher also observed that low-level students in homogeneous groups easily went off-task, and students in one heterogeneous group could not get along. As suggested by Chen (2012) and Zhan and Chen (2012), more facilitation should be provided to lower achievers. This study shows that dysfunctional groups need immediate help. In the future, the teacher would rearrange group members as soon as a problem became evident. Also the teacher concluded that homogeneous grouping of low achievers should be avoided and grouping needs to be more flexible, taking students' characteristics into account. In addition to language levels, the teacher would like to try mixing extroverted with introverted students to see if balancing their different tendencies could improve collaboration.

Additionally, students' group work took a lot more class time than was initially allocated, probably because students were not used to collaborative work and had to adjust to working with others. DI and collaborative learning originated from Western L1 contexts (Vygotsky, 1978; Washburne, 1953). As Bowers (2005) suggests, Western theories should not be idealized in Eastern contexts nor normalized to local situations. Teachers need to work as cultural mediators to provide explicit instruction in collaborative work, such as how to express one's opinions and accept others' ideas, so students can successfully cooperate.

Individual task choice was used as the wrap-up evaluation. This activity was highly welcomed by the students. Being offered choices enabled students to start from their current levels and challenge themselves to achieve at higher levels. By doing so, students enjoyed learning autonomy and took ownership of their learning. Most importantly, the process brought them a sense of achievement; providing even more levels of each task is recommended. Not only the mid and low level students, as in this study, but higher level students should also have opportunities to move up to different levels. This suggestion echoes Krashen's (1985) idea of i+1. Students learn better when the content is slightly higher than their current level. In the DI classroom, Krashen's idea of i+1 and Vygotsky's ZPD are complementary in emphasizing the effectiveness of starting with what learners have mastered and guiding them to take reasonable steps beyond to incrementally higher levels of mastery at a pace that is manageable and at the same time reasonably challenging to the learner.

This study is at an initial stage of DI research and has limitations that recommend it as a pilot study for further investigation. The first limitation was the small number of participants, 26 students divided into three levels, so results cannot be generalized beyond the context of the study. Second, the young learners may not have been able to express themselves well in the survey. Various types of qualitative data, such as interview and class observations, will enrich understanding of their experiences and provide a detailed picture of the entire process.

This study also calls for future research in several ways. First, the course design in this study focuses on a multi-level classroom. Courses that differentiate other perspectives such as students' learning style preferences can be explored. Second, homogeneous grouping and heterogeneous grouping can be compared to reveal which approach is appropriate in what contexts and why. Third, students' opinions were elicited with only a three-point scale survey due to the young age of the participants. More qualitative data, such as class observation and interviews, would be useful to uncover students' attitudes towards each activity. Fourth, inquiry involving quantitative data and more sophisticated analyses should be conducted with larger groups of participants.

In short, a teacher seeks to achieve a balance between teaching a full-sized class and addressing individual needs in DI. It provides a way to anticipate the needs of students in small groups characterized by similar levels of present mastery. Although it may be labor-intensive at the beginning, over time, a teacher can develop a useful repertoire of alternative tasks while becoming more skilled at assessing varying levels. The way that DI addresses the conceptual framework of socio-cultural theory and i+1 is that the student actively participates in deciding which level of tasks to try. This feature of student choice makes DI different from tracking. In this way students can show the teacher where they want to begin and, bolstered by success, make their own moves to higher levels. This collaboration between teacher and student in forging the student's path is one of the key elements in DI, which is a way to plausibly address the problem of meeting the needs of students at varying levels in a single classroom.

References

Alavinia, P., & Farhady, S. (2012). Using differentiated instruction to teach vocabulary in mixed ability classes with a focus on multiple intelligences and learning styles. *International Journal of Applied Science and Technology, 2*(4), 72-82.

Alavinia, P., & Sadeghi, T. (2013). The impact of differentiated task-based instruction via heeding learning styles on EFL learners' feasible proficiency gains. 3L: *The Southeast Asian Journal of English Language Studies, 19*(1), 75-91.

Ankrum, J. W., & Bean, R. M. (2008). Differentiated reading instruction: What and how. *Reading Horizons, 48*, 133-146.

Beal, K. (1991). *I love my family.* Boston, MA: Addison-Wesley Publishing Company.

Bowers, C. A. (2005). *The false promises of constructivist theories of learning.* New York, NY: Peter Lang Publishing, Inc.

Chen, S., & Tsai, Y. (2012). Research on English teaching and learning: Taiwan (2004-2009). *Language Teaching, 45*, 180-201.

Chen, Y. H. (2007). *Exploring the assessment aspect of differentiated instruction: College EFL learners' perspectives on tiered performance tasks.* Unpublished doctorate dissertation, University of New Orleans.

Chern, C. (2002). English language teaching in Taiwan today. *Asia-Pacific Journal of Education, 22,* 97-105.

Chien, C.-W. (2012). Differentiated instruction in an elementary school EFL classroom. *TESOL Journal, 3,* 280-291.

Coulter, S. E., & Groenke, S. L. (2008). A differentiated vocabulary unit for John Knowles's A Separate Peace. *English Journal, 97*(4), 26-32.

Dahlman, A., Hoffman, P., & Brauhn, S. (2008). Classroom strategies and tools for differentiating instruction in the ESL classroom. *MinneWITESOL Journal, 25*, 58-77.

Gregory, G., & Chapman, C. (2002). *Differentiated instructional strategies: One size doesn't fit all.* Thousand Oaks, CA: Corwin Press.

Haager, D., & Klingner, J. K. (2005). *Differentiating instruction in inclusive classrooms.* Columbus, OH: Merrill.

Halliday, M. A. K. (1993). Towards a language-based theory of learning. *Linguistics and Education, 5,* 93-116.

Heacox, D. (2002). *Differentiating instruction in the regular classroom: How to reach and teach all learners, grades 3-12.* Minneapolis, MN: Free Spirit.

Kaufman, D. (2004). Constructivist issues in language learning and teaching. *Annual Review of Applied Linguistics, 24,* 303-319.

King-Shaver, B., & Hunger, A. (2003). *Differentiated instruction in the English classroom: Content, process, product, and assessment.* Portsmouth, UK: Heinemann.

Krashen, S. (1985). *The Input Hypothesis: Issues and implications.* New York, NY: Longman.

Lawrence-Brown, D. (2004). Differentiated instruction: Inclusive strategies for standards-based learning that benefit the whole class. *American Secondary Education, 32*(3), 34-62.
Levy, H. M. (2008). Meeting the needs of all students through differentiated instruction: Helping every child reach and exceed standards. *The Clearing House, 81,* 161-164.
Lewis, S., & Batts, K. (2005). How to implement differentiated instruction? Adjust, adjust, adjust. *Journal of Staff Development, 26*(4), 26-31.
Liu, Y. F. (2008). Differentiated instruction in flexible grouping in EFL classrooms. *Journal of Taipei Municipal University of Education, 39*(1), 97-122.
Magno, C., & Sison, K. G. (2011). Reading English words aloud: Will it help or will it not? *TESOL Journal, 4*, 64-71.
Mastropieri, M. A., Scruggs, T. E., Norland, J. J., Berkely, S., McDuffie, K., Tornquist, E. H., & Connors, N. (2006). Differentiated curriculum enhancement in inclusive middle school science: Effects on classroom and high-states tests. *The Journal of Special Education, 40*(3), 130-137.
Nordlund, M. (2003). *Differentiated instruction: Meeting the educational needs of all students in your classroom.* Lanham, MD: Scarecrow Education.
Odgers, S., Symons, A., & Mitchell, I. (2000). Differentiating the curriculum through the use of problem solving. *Research in Science Education, 30*, 289-300.
Palinscar, A. S. (2005). Social constructivist perspectives on teaching and learning. In H. Daniels (Ed.), *An introduction to Vygotsky* (2 ed., pp. 285-314). New York, NY: Routledge.
Rock, M. L., Gregg, M., Ellis, E., & Gable, R. A. (2008). REACG: A framework for differentiating classroom instruction. *Preventing School Failure, 52*(2), 31-47.
Schumm, J. S., Moody, S. W., & Vaughn, S. (2000). Grouping for reading instruction: Does one size fit all? *Journal of Learning Disabilities, 33*, 477-488.
Sivasubramaniam, S. (2011). Constructivism in EIL: Issues and insights for teaching and research. *The Journal of English as an International Language, 1*, 1-20.
Shen, C., Lin, F., Xu, Y., Guo, W., Guo, F., Chen, M., Liu, S. (2001). Enjoy 2. Tainan, Taiwan: Bureau of Education of Tainan City Government.
Subban, P. (2006). Differentiated instruction: A research basis. *International Education Journal, 7,* 935-947.
Tieso, C. (2003). Ability grouping is not just tracking anymore. *Proper Review, 26*, 29-36.
Tieso, C. (2005). The effects of grouping practices and curricular adjustments on achievement. *Journal for the Education of the Gifted, 29*, 60-89.
Tomlinson, C. A. (1999). *The differentiated classroom: Responding to the needs of all learners.* Alexandria, VA: Association for Supervision and Curriculum Development.
Tomlinson, C. A. (2004). Sharing responsibility for differentiating instruction. *Proper Review, 26*, 188.
Tomlinson, C. A. (2005). *How to differentiate instruction in mixed-ability classrooms (2nd ed.).* Upper Saddle River, NJ: Pearson, Merrill Prentice Hall.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: MIT Press.
Vygotsky, L. S. (1981). The instrumental method in psychology. In J. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 3-35). Armonk, NY: Sharpe.
Washburne, C. W. (1953). Adjusting the program to the child. *Educational Leadership, 11*, 138-147.
Wells, G. (2007). Semiotic mediation, dialogue and the construction of knowledge. *Human Development, 50*, 244-274.
Wertsch, J. (1991). *Voices of the mind: A sociocultural approach to mediated action.* Cambridge, MA: Harvard Univ. Press.
Zhan, Y., & Chen, J. (2012). The effects of balanced reading instruction with differentiated grouping on elementary students' English word recognition, reading comprehension, and reading attitudes. *English Teaching & Learning, 36*(2), 125-163.

The Effect of Text Enhancement on Learning and Recall of English Language Structures And Its Interaction with Gender and Personality-Traits

Mohammad Tajbakhsh*
Iran Ministry of Education, Kahrizak Education District, Iran

Nasser Gharakhanloo
Kharazmi University, Tehran, Iran

Abstract

This study employed a pretest-posttest control group design supplemented by a delayed posttest to compare the effects of textually enhanced grammar instruction with traditional explicit grammar instruction in an experimental and a control group. Since it was perceived that individual differences as contributing factors were not included in previous studies, the effect of students' gender and personality traits were also investigated. The results of the study revealed that explicit grammar instruction, compared to Textual Enhancement (TE), promoted grammar achievement after a month of instruction; however, such superiority did not last over a longer period. Moreover, TE contributed to better recall of grammar structures among female EFL learners. Regardless of teaching methods, high Openness and high Conscientiousness learners had better achievement of structures, although only high Conscientiousness could maintain its supremacy after two weeks. The results of this investigation provide further evidence to the current debate about the true potentials and limitations of TE, and contribute to a comprehensive model of meaningful grammar instruction.

Keywords: focus on form, grammar teaching, attention, EFL learners

Introduction

The current consensus among language practitioners is that language learning should have a primary focus on meaning within an overall communicative framework (Ellis, 2006). However, such a framework fails to strike a balance between communicative fluency and formal accuracy. Studies (e.g., Long, 1996; Long & Robinson, 1998) have shown that students' language learning, void of any grammar instruction, would not lead to native-like grammatical competence. Thus, a reconsideration of grammar teaching gave rise to a focus on forms (FonfS) and focus on form (FonF) dichotomy. The former advocates a specific structure to be taught explicitly where language is treated primarily as an object to be learned and practiced through a synthetic syllabus. On the contrary, FonF is indirect and context-based in which meaning making rather than explicit teacher-led instruction of grammar is the core objective (Doughty & Williams, 1998). The latter method seems to be more suited to the current state of foreign language teaching.

Ellis (2001) categorizes FonF into two types, Incidental and Planned. Incidental FonF assumes that classroom activity should be based on communicative tasks and that attention to form should be paid only when grammatical difficulties arise which lead to (or anticipate) a communicative breakdown (Bernard & Scampton, 2008). Planned FonF, on the other hand, involves the treatment of predetermined grammatical features, but it

* *Corresponding author. Tel: 98-9124595368; Email:* MohammadTajbakhsh@yahoo.com*; Address: ELT Group, Kahrizak Education District Office, Shahid Mostafavi Street, Kahrizak, Tehran, Iran. Postal Code: 1816148581*

differs from FonfS as it occurs when the learners' attention is primarily engaged in meaning making processes through the use of focused tasks in an analytical syllabus (Bernard & Scampton, 2008). Input Enhancement is one of the techniques in planned FonF teaching. The aim of this technique is to increase the attention learners' pay to linguistic form by making target structure more salient through various typographical devices such as bolding, underlining, and italicizing in the written input. This technique is not limited to written input; oral input, for instance, can also be enhanced by some acoustic devices such as added stress or repetition. Moreover, in the target language input, those language components to which learners pay attention are believed to be learned better (e.g., Leeman, Arteogotia, Fridman, & Doughty, 1995; Leow, 1997; Rosa & O Neill, 1999), and input enhancement aims to facilitate such attention.

Initially proposed by Sharwood Smith (1981) as consciousness-raising, input enhancement (Sharwood Smith, 1991, 1993) was an attempt to solve the disadvantages of analytic approaches towards language through augmenting them with synthetic techniques; that is, making input more noticeable in all aspects of the target language, such as lexicon, phonology, grammatical form, and pragmatics (Jin, 2011).

Although such attention has a crucial role in mediating input and learning (e.g., Robinson, 1995; Schmidt, 1990, 1993, 1994, 1995, 2001; Tomlin & Villa, 1994), there is still disagreement about the amount of attention expended by learners. Moreover, the amount and level of input (Gass & Selinker, 2008) are not solely sufficient to influence intake out of input. The key to such a process is noticing (Gass, 1997; VanPatten, 1996). Schmidt (1994) signified that the subliminal perception of input does not lead to foreign language learning; rather, input should be consciously noticed by learners. Jin (2011) also tried to examine the role of consciousness in second language learning and concluded that language learning is explicit. They believed that incidental learning is both possible and effective if the nature of the task calls for attention to what is being learned. That is to say, task based language teaching (mainly implicit meaningful processes) and Schmidt's Noticing Hypothesis (Schmidt, 1990, 1993, 1994, 1995, 2001) (explicitly attended materials) could be implemented simultaneously. On the other hand, it is clear that some of our learning does not take place consciously of our own volition. What we learn throughout our lives is not bound to learning in classroom environments or other sorts of instruction we might receive. We improve our knowledge and, in general, cognitive abilities in ways other than through formal instruction. This kind of learning plays an important role in our skills, perceptions, and behaviors. Kaufman et al. (2010) characterized this learning by "a set of automatic, associative, nonconscious, and unintentional learning processes as distinguished from the conscious, deliberate, and reflective learning processes" (p. 321). Considering Kaufman et al. (2010) and Jin (2011), the balance between explicit and intentional, and implicit and unintentional ways of learning could open up new possibilities into boosting learning.

Similarly, formal grammar instruction in a classroom can be augmented through meaningful language instruction. Focus on Form is an effort to create such a middle ground for form instruction in overall language pedagogy. An input enhancement technique to enhance noticing of input in planned FonF is called Textual Enhancement (TE). The main idea behind TE is to draw learners' attention to linguistic features by modifying the physical appearance of written text through modifications such as underlining, boldfacing, italicization, capitalization, color coding, or using different font sizes or types. These are some ways of improving the perceptual saliency of form in second language input; and consequently, making it more noticeable and easier to process.

In the last few decades, the large number of studies on the effects of TE on grammar learning has revealed more controversies than conclusions. Shook (1994) rendered positive effects for TE on the intake of the present perfect and the relative pronoun que/quien by L2 Spanish learners. Similarly, Doughty (1991) and Leeman, Arteogotia, Fridman, and Doughty (1995) showed that TE has positive effects over the use of explicit rule explanation with adult learners.

In another study, Alanen (1995) examined the effects of TE and explicit rule presentation on 36 Finish language learners who were divided into four groups. Group A received a plain text, with no modification, group B received enhanced text, group C received grammar rule explanation, and group D received both grammar rule explanation and enhanced text. It was found that although groups exposed to TE performed better than the control group, their structure achievement was lower than those of groups C and D.

In the same fashion, Pacheco (2004), in a quasi-experimental classroom study, put the effectiveness of the textual enhancement method under investigation and compared it with explicit rule-based teaching. Some grammatical features of English grammar known to have been problematic for young L2 learners were chosen and the effects of text enhancement and explicit rule-based teaching in a four-week period on four groups of learners were compared. The performances of these four groups can be represented as R+>R>E>C with group (R+) receiving both explicit rule explanation and visually enhanced materials, (R) explicit rule explanation, (E) enhanced texts, and finally (C) without rule explanation or visual enhancement. In comparison to similar studies, Pacheco's (2004) findings may prove the moderating effect of the age of participants on the effects of TE.

Abe (2006) found that interactive input enhancement had a positive effect on the acquisition of connected speech by Japanese college students, while Lee (2007), using a correction task to measure intake, and a free recall task to measure comprehension, found that TE contributed to the acquisition of targeted forms, but negatively affected the comprehension of Korean adult ESL learners.

De Santis (2008) also led research on L2 learners who were exposed to target language only in aural mode. The effects of TE were investigated on adult L2 learners who had been exposed to the target forms via naturalistic aural input. In this study, the groups who received TE materials outperformed the control group in terms of their awareness of target forms; however, such treatment appeared to have no facilitative effect on their oral production.

More recently, TE has also been studied in the Iranian context. Vaez Dalili, Ketabi, Kassaian, and Eslami Rasekh (2011) compared the effects of textual enhancement and explicit metalinguistic instruction on the learning of English dative alternation among Iranian lower-intermediate EFL learners and showed that the group receiving explicit metalinguistic instruction outperformed the group which received TE instruction.

Moreover, the impact of TE on the acquisition of English conditional sentences was investigated by Rashtchi and Gharanli (2010). The results showed that the performance of the TE group was much higher than that of the control group which was instructed through explicit grammar teaching.

In another study, Sarkhosh and Sarboland (2012) put different textual enhancements under investigation. Results revealed underlined and bold were more effective than other TE formats. The performance of the group which received underlined material showed much better results. In another study, Farahani and Sarkhosh (2012) came to the same conclusion.

The high number of studies on the benefits of the synergy between meaning oriented tasks and learners' noticing of formal aspects of language seems to indicate the controversial nature of these issues, and more investigations would undoubtedly lead to a more comprehensive knowledge of their potential. In the same fashion, FonF and TE, among the options in an EFL teacher's set of possible techniques for implementing meaningful grammar instruction, are in need of deeper investigation for their potential to be realized in different contexts.

However, none of the previous investigations, to the researchers' knowledge, included the effects of the commonest individual differences (IDs) such as gender and/or personality types (cf. Rashtchi & Gharanli, 2010; Vaez et al., 2011) on learners' achievement, or more significantly on their recall of grammatical structures over a longer period (cf. Farahani & Sarkhosh, 2012; Rashtchi & Gharanli, 2010; Sarkhosh & Sarboland, 2012;). This is why these two commonly investigated IDs were included in this study.

Learners' Individual Differences

As mentioned earlier, attention is considered to be the sole effective factor to learning (e.g., Leeman, et al., 1995; Leow, 1997; Rosa & O Neill, 1999). However, Park (2004) and De Santis (2008) concluded that multiple individual factors also contribute to learners' attention to and benefit from enhanced materials.

The deep-rooted issue of the effects of IDs in learning, especially how each student may benefit from explicit and/or implicit sides of instruction, induces disparities in every new context.

One such ID is personality. Personality theory assumes that individuals are characterized by nearly unchanging patterns of traits. One well-known example of these hierarchical models is the Big Five Trait taxonomy (Digman, 1990; Goldberg, 1981; McCrae & Costa, 1990). This model consists of five factors, namely: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. The difference between this theoretical framework and previously well-known templates such as the Myers–Briggs Type Indicator (MBTI) (Myers, 1962) lies in that the concept of *trait*, "denote[s] consistent patterns of behavior, especially expressive or stylistic behavior" (John, Robins, & Pervin, 2008, p.11) while *personality type* delineates "coherently organized bundles of trait-like characteristics that define interesting patterns" (John et al., 2008, p.12).

The "Big Five" personality dimensions do not nullify other taxonomies and personality conceptualizations, and are derived from the terms which people have used to describe themselves and others. In fact, it represents, and could contribute to other personality descriptions (John et al., 2008, p. 116). Table 1 shows personality trait domains in this model.

Learners' gender, another so commonly-debated issue, is a biological factor affecting language learning. The literature about gender shows that it does make differences in language learning, but findings of the effect of gender difference on learning are not uniform. The discrepancies in the existing literature on the relationship of personality traits and gender with language learning have pushed them to a very relative and mundane level, suggesting that any attempt to provide a conclusive account of such scattered findings has no benefit. In other words, the consensus of studies on gender and/or personality effects on second/foreign language learning cannot be reached without including so many other contextual factors, which goes beyond the extent of any study. Yet, the very omnipresence of these factors makes their inclusion in any study nearly inevitable. Moreover, the results of previous studies on the effects of TE on EFL learners' grammar acquisition are far from unified. Some studies, especially in the Iranian context, report inappropriate methodology for participant sampling (Farahani & Sarkhosh, 2012; Sarkhosh & Sarboland, 2012) and nearly all of the previous investigations were focused on just one grammatical point in the treatment rather than teaching and assessing a variety of grammatical topics during a course of instruction. Hence, this study tries to compare the effects of TE and traditional explicit rule explanation on EFL learners' achievement and recall of English structures, and their interactions with personality traits and gender through posing the following research questions:

1. Are there any significant differences between the textual enhancement method and an explicit grammar instruction method in terms of their effects on learners' achievement of language structures?
2. Are there any significant differences between the textual enhancement method and an explicit grammar instruction method in terms of their effects on learners' recall of language structures?
3. Does gender make a difference in the effects of the two methods on the achievement of language structures?
4. Does gender make a difference in the effect of the two methods on the recall of language structures?
5. Do personality traits make a difference in the effects of the two methods on the achievement of language structures?
6. Do personality traits make a difference in the effects of the two methods on the recall of language structures?

Table 1
The Ocean of Personality: Definition and Explication of the Big Five Domains (John et al., 2008, p.120)

Factor initial (number)	E (Factor I)	E (Factor II)	E (Factor III)	E (Factor IV)	E (Factor V)
Verbal labels	Extraversion Energy Enthusiasm	Agreeableness Altruism Affection	Conscientiousness Constraint Control of impulse	Neuroticism Negative Emotionality Nervousness	Openness Originality Open-Mindedness
Conceptual definition	Implies an *energetic approach* toward the social and material world and includes traits such as sociability, activity, assertiveness, and positive emotionality.	Contrasts a *prosocial and communal orientation* toward others with antagonism and includes traits such as altruism, tender-mindedness, trust, and modesty.	Describes *socially prescribed impulse control* that facilitates task- and goal directed behavior, such as thinking before acting, delaying gratification, following norms and rules, and planning, organizing, and prioritizing tasks.	Contrasts emotional stability and even-temperedness with *negative emotionality*, such as feeling anxious, nervous, sad, and tense.	Describes the breadth, depth, originality, and complexity of an individual's *mental and experiential life*.
Behavioral examples	Approach strangers at a party and introduce myself; Take the lead in organizing a project; Keep quiet when I disagree with others (R)	Emphasize the good qualities of other people when I talk about them; Lend things to people I know (e.g., class notes, books, milk); Console a friend who is upset	Arrive early or on time for appointments; Study hard in order to get the highest grade in class; Double-check a term paper for typing and spelling errors; Let dirty dishes stack up for more than one day (R)	Accept the good and the bad in my life without complaining or bragging (R); Get upset when somebody is angry with me; Take it easy and relax (R)	Take the time to learn something simply for the joy of learning; Watch documentaries or educational TV; Come up with novel setups for my living space; Look for stimulating activities that break up my routine
Examples of external criteria predicted	*High pole:* Social status in groups and leadership positions; selection as jury foreperson; positive emotion expression; number of friends and sex partners *Low pole:* Poorer relationships with parents; rejection by peers	*High pole:* Better performance in work groups *Low pole:* Risk for cardiovascular disease, juvenile delinquency, interpersonal problems	*High pole:* Higher academic grade-point averages; better job performance; adherence to their treatment regimens; longer lives *Low pole:* Smoking, substance abuse, and poor diet and exercise habits; attention-deficit/hyperactivity disorder (ADHD)	*High pole:* Poorer coping and reactions to illness; experience of burnout and job changes *Low pole:* Feeling committed to work organizations; greater relationship satisfaction	*High pole:* Years of education completed; better performance on creativity tests; success in artistic jobs; create distinctive-looking work and home environments *Low pole:* Conservative attitudes and political party preferences

Note. Conceptual definitions are based on John and Srivastava (1999). Behavioral examples are based on significant correlations between Big Five Inventory scales and self-reported act frequencies in an undergraduate sample (N = 375; John & Naumann, 2007). (R) denotes that the act was a reverse-keyed item (i.e., correlated negatively with the Big Five domain). The first letters of the Big Five dimensions form the anagram OCEAN.

Method

Participants

From among the EFL students of two English institutes in Saveh, four classes of an intermediate level were chosen to participate in this study. (Saveh is located in central Iran, in Markazi province, and has a population of about 180,000.) Each of these institutes administers single-sex EFL classes for both genders, and from each institute one male and one female class were chosen. In total, four classes with 17 learners in each participated. Eight students were excluded due to their inhomogeneous ages and performances on the proficiency test. Subsequently, 60 participants in two classes of males and two classes of females were included in the analysis after administering the Oxford proficiency test.

All participants were young adults with males ranging from 11 to 16 (M= 14) and females from 10 to 17 (M= 14).

Instruments

Oxford Quick Placement Test (OPT, V 1)

This test contains 60 items. It tests English learners' proficiency using grammatical items over 30 minutes. The test takers' proficiency is measured through their scores ranging from Beginners to Upper advanced as follows:

1-17 (Beginner)
18-27 (Elementary)
28-36 (Lower intermediate)
37-47 (Upper intermediate)
48-55 (Advanced)
56-60 (Upper advanced)

Since the above scale does not contain any intermediate level, scores from 32 to 42 were considered as intermediate level. The OPT achieved an acceptable reliability (α =.84) in this study.

The Big Five Inventory (BFI)

The BFI ((John, Donahue, & Kentle, 1991) contains 44 items on a 5 point Likert scale which has been proven to be a more reliable instrument to measure personality traits in comparison to older personality trait questionnaires (John et al., 2008). BFI proved to be a psychometrically appropriate personality scale (John et al., 2008) in comparison to Trait Descriptive Adjectives (TDA) (Goldberg, 1992) and the NEO Five-Factor Inventory(NEO-FFI) (Costa & McCrae, 1992a, 1992b). Table 2 shows the reliability and validity comparison of these inventories.

As Table 2 (reproduced from John et al., 2008, p.132) shows, BFI scored acceptable alpha reliability mainly in the North American context. The BFI achieved a Cronbach's alpha reliability of .81 from the 60 participants in this study.

Table 2
Reliability and Convergent Validity Coefficients for Three Short Big Five. Measures: Big Five Inventory, NEO Five-Factor Inventory, and Trait Descriptive Adjectives

Measures	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness	Mean
			Internal consistency			
BFI	.86	.79	.82	. 87	.83	.83
NEO-FFI	.82	.75	.82	.87	.76	.81
TDA	.88	.84	.84	.83	.83	.84
Mean	.85	.80	.83	.85	.81	.83
			Uncorrected convergent validity correlations (across measures)			
BFI–TDA	.90	.75	.79	.70	.79	.80
BFI–NEO-FFI	.73	.76	.80	.81	.72	.77
TDA–NEO-FFI	.70	.66	.75	.64	.62	.68
Mean	.80	.73	.78	.73	.72	.75
			Corrected convergent validity correlations (across measures)			
BFI–TDA	.99	.93	.96	.82	.95	.95
BFI–NEO-FFI	.87	.99	.97	.94	.90	.95
TDA–NEO-FFI	.83	.83	.91	.76	.78	.83
Mean	.94	.95	.95	.86	.90	.93
			Standardized convergent validity coefficients from CFA (controlling for acquiescence factors)			
BFI–TDA	.99	.91	.91	.84	.97	.95
BFI–NEO-FFI	.83	.98	.95	.93	.90	.93
TDA–NEO-FFI	.76	.84	.87	.78	.74	.80
Mean	.92	.93	.92	.86	.91	.91

Text Enhanced Materials
According to Sharwood Smith (1993), it is possible to highlight either correct or incorrect form. They called it positive input enhancement when correct form is highlighted and negative enhancement when erroneous form is highlighted. The visual enhancement which was utilized in this study is of the positive input enhancement type. Grammatical topics taught during the study were Participle Adjectives, Relative clauses, Modality, Conditional type 1, Perfect Modals, and Reported Speech.

For the purpose of positive enhancement, the grammatical structures available in the learners' course book, New Interchange 3 (Richards, Hull, & Proctor, 2005), were chosen, redesigned, and enhanced textually. These grammatical topics were taught in separate sessions and took about 45 minutes each. The TE formats used in this study were the following two types: boldfaced and underlining. (Appendix)

To design the tests for the purpose of this study, the grammar tests of the English institutes were used. The test contained all of the grammatical topics mentioned above. Eighty test items were chosen by extracting items based on the grammatical structures taught in the courses. To have two parallel 40-item test versions with homogenously distributed items of all the grammar structures, questions with odd numbers were used in one version and even numbers in the other. The pretest, posttest, and delayed posttest versions achieved acceptable reliabilities ranging from .66 to .71 which were considered acceptable.

Pretest and Posttest

For pretest and posttest, the first 40-item test was administered twice with a four-week period in between.

Delayed Posttest

Two weeks after the posttest, the second 40-item version was used as the delayed posttest to assess the students' recall of taught structures.

Procedure

The present study was a pretest-posttest control group design (Figure 1). The participants were already assigned their classes with each of the above-mentioned English institutes having one male and one female class. After assigning the experimental and control groups, a proficiency test (PT) was administered to confirm the participants' homogeneity in terms of language proficiency. Furthermore, a pretest of grammar (T1) was given to both groups. While the experimental group was receiving the TE treatment (X) for 8 sessions during the term, the control group received the explicit explanation grammar instruction. Then the very same test was given to both groups as the posttest of grammar (T2) to compare the gains of participants in each group in their use of the grammar structure. Finally, after two weeks, a delayed posttest (T3) was administered to measure the participants' recall of grammatical topics covered during the course. Moreover, the personality traits questionnaire was given to the participants at the end of the course.

G1 (experimental)	PT	T1	X	T2	T3	BFI
G2 (control)	PT	T1		T2	T3	BFI
	Week 1			Week 5	Week 7	

Figure 1. Study Design

Each group consisted of 30 participants, 15 males and 15 females in each class. One of the researchers and a comprehensively briefed female colleague taught the experimental group classes, whereas two other teachers were selected to teach the control classes. The teachers in the control classes were requested to teach grammar topics using the explicit explanation instruction method.

It should be mentioned that students in the experimental group were very curious about the new style and typographical differences of the materials provided to them. This curiosity was shown through their posing of many questions; however, the teachers did not provide any explanation about these differences; rather, they provided feedback such as "compare and contrast the bold faced with underlined terms", "try to come up with rules". No meta-textual explicit explanation of grammar rules was provided to students. Moreover, the researchers were present in all the cases of test administration to provide help and instructions to participants.

Results

To investigate the difference in the performances of EFL learners in two groups of explicit grammar instruction (EXGI) and textual enhancement grammar instruction (TEGI) on the posttest and the delayed posttest two independent T-tests were run.

Effects of Teaching Methods on Achievement of Language Structures

The first independent T-test showed a significant difference t (58) =2.52, p<.05 (Table 3) between the performance of EXGI and TEGI groups on the posttest of grammar as the learners in EXGI (M=28.8, SE=.55) outperformed those in TEGI (M= 26.9, SE=.51).

Table 3
Independent Samples Test for EXGI and TEGI on the Posttest

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Posttest	Equal variances assumed	.003	.957	2.523	58	.014	1.90000	.75315	.39240	3.40760
	Equal variances not assumed			2.523	57.620	.014	1.90000	.75315	.39219	3.40781

Effects of Teaching Methods on Recall of Language Structures

However, the second independent T-test analysis showed no significant difference t (58) =1.24, $p>.05$ between the performance of EXGI (M=22.4, SE=.45) and that of TEGI (M=23.2, SE=.48) in the recall of the language structures in the delayed posttest.

Effects of Learners' Gender on Achievement and Recall of Language Structures

To probe the effect of learners' gender on achievement of language structures in the posttest, a two-way ANOVA was applied. The same analysis was also utilized to investigate the effect of gender on recall of language structures in the delayed posttest. The first revealed no significant effect of learners' gender on achievement $F(1, 56)=2.14$, $p>.05$, $h_p^2=.037$ in the posttest. However, the second indicated a significant effect $F(1, 56)=7.56$, $p<.05$, $h_p^2= .12$ (Table 4) of learners' gender on recall; as females (M=23.7, SE=.45) performed better than males (M=21.9, SE=.45) in the delayed posttest.

The interaction between learners' gender and teaching methods revealed no significant effect on either the achievement of language structures in the posttest $F(1, 56)=15$, $p>.05$, $h_p^2= .003$), or on the recall of them in the delayed posttest $F(1, 56)=.33$, $p>.05$, $h_p^2= .006$.

Table 4
Tests of Between-Subjects Effects for Gender, Teaching Methods, and Their Interaction on the Delayed Posttest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Teaching Methodology	10.417	1	10.417	1.704	.197	.030
Gender	46.817	1	46.817	7.657	.008	.120
Teaching Methodology * Gender	2.017	1	2.017	.330	.568	.006
Error	342.400	56	6.114			
Total	31729.000	60				

Note. R Squared = .148 (Adjusted R Squared = .102)

Effects of Personality Traits on Achievement of Language Structures

Five separate two-way ANOVAs were also run to probe any significant effect of learners' personality traits on methods of grammar teaching in the posttest.

Learners' openness showed a significant effect F (1, 56) = 4.85, p<.05, h_p^2=.8 (Table 5) on achievement in the posttest. The mean scores for the high Openness and low Openness learners were M=28.8 (SE=.48) and M=26.9 (SE=.55), respectively.

The interaction between learners' Openness and teaching methods revealed no significant effects F (1, 56) =1.23, p>.05, h_p^2=.021 on achievement of the structures in the posttest.

Table 5
Tests of Between-Subjects Effects for Openness, Teaching Methods, and Their Interaction on the Posttest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Teaching methodology	52.890	1	52.890	6.664	.012	.106
Openness	38.551	1	38.551	4.857	.032	.080
Teaching Methodology * Openness	9.762	1	9.762	1.230	.272	.021
Error	444.465	56	7.937			
Total	47085.000	60				

Note. R Squared = .188 (Adjusted R Squared = .145)

Conscientiousness appeared to have a significant effect F (1, 56) = 8.6, p<.05, h_p^2=.13 (Table 6) on achievement with the mean score of high Conscientiousness 28.7 (SE=.47) being well above that of low Conscientiousness learners 26.6 (SE=.54).

On the other hand, Conscientiousness interacting with teaching methods revealed no significant effects F (1, 56) =.47, p>.05, h_p^2=.001 on achievement of structures in the posttest.

However, the two-way ANOVAs showed no significant effects for Extraversion F (1, 56) =.27, p>.05, h_p^2=.005, Agreeableness F (1, 56) =.005, p>.05, h_p^2=.0, or Neuroticism F (1, 56) =.29, p>.05, h_p^2=.005 on achievement in the posttest. Similarly, the interaction of each of these personality traits with teaching methods revealed no significant effects on learners' performance on the posttest.

Table 6
Tests of Between-Subjects Effects for Conscientiousness, Teaching Methods, and Their Interaction on the Posttest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Teaching Methodology	54.359	1	54.359	7.121	.010	.113
Conscientiousness	65.648	1	65.648	8.600	.005	.133
Teaching Methodology * Conscientiousness	.359	1	.359	.047	.829	.001
Error	427.493	56	7.634			
Total	47085.000	60				

Note. R Squared = .219 (Adjusted R Squared = .178)

Effects of Personality Traits on Recall of Language Structures

Five separate two-way ANOVAs were run to investigate any significant effects of personality traits recall of language structures in the delayed posttest.

All personality traits, except for Conscientiousness F (1, 56) = 6.89, p<.05, h_p^2=.11 (Table 7), revealed no significant effects on recall of language structures in the delayed posttest; Openness F (1, 56) =.009, p>.05, h_p^2 =.000, Extraversion F (1, 56) = .99, p>.32, h_p^2 =.018, Agreeableness F (1, 56) =.58, p>.32, h_p^2=.01, and Neuroticism F (1, 56) =.2, p>.05, h_p^2=.016.

The mean scores for high Conscientiousness and low Conscientiousness were 23.5 (SE=.42) and

21.8 (SE=.48), respectively. Moreover, neither the interaction of Conscientiousness F (1, 56) =.19, p>.05, h_p^2 =.003, or the other personality traits with teaching methods showed any significant effects on the learners' recall of language structures.

Table7
Tests of Between-Subjects Effects for Conscientiousness, Teaching Methods, and Their Interaction on the Delayed Posttest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Teaching Methodology	11.178	1	11.178	1.802	.185	.031
Conscientiousness	42.761	1	42.761	6.895	.011	.110
Teaching Methodology * Conscientiousness	1.178	1	1.178	.190	.665	.003
Error	347.294	56	6.202			
Total	31729.000	60				

Note. R Squared = .135 (Adjusted R Squared = .089)

Discussion

With a pretest-posttest control group design supplemented by a preliminary general proficiency test and a delayed posttest, this study investigated the effects of textual enhancement grammar teaching in comparison to traditional explicit grammar teaching on students' achievement and recall of grammatical structures. Results revealed a significant difference in the grammar achievement of learners in the posttest as the result of teaching methods. Students in the explicit grammar explanation groups performed better than those who received TE materials. This, concurs with Leow (1997), Overstreet (1998), Wong (2000), Leow, Egi, Nuevo, and Tsai (2003), Radwan (2005), and contradicts Doughty (1991), Shook (1994), Leeman et al. (1995), Jourdenais, Ota, Stauffer, Boyson, and Doughty (1995), Alanen (1995), and Abe (2006), showing that teaching grammar through textually enhanced materials leads to lower levels of grammar acquisition among intermediate EFL learners compared to explicit metalinguistic explanation.

Moreover, results on the learners' recall of language structures showed no significant differences due to different methods of grammar instruction. Indeed, the significant effects of the explicit grammar instruction seemed to disappear after a two-week period. In other words, it is safe to claim that the supremacy of the traditional explicit method cannot be maintained over longer periods.

The instructional method used, either explicit grammar instruction or textual enhancement instruction, showed no effects based on a learner's gender. Male and female learners acquired structure equally; however, in the delayed posttest, females retained taught structures much better than males.

This investigation of the effects of personality traits on teaching methods, in line with (Kaufman et al., 2010), showed that no matter which teaching method was used, high Openness and high Conscientiousness learners were better users of grammar instruction than low Openness and low Conscientiousness learners, respectively. High Conscientiousness learners, were found to recall language structures much better than low Conscientiousness and other personality traits. High Conscientiousness also includes following rules and thinking before acting, in fact, high Conscientiousness learners seem to be better rule observers no matter what and how grammatical points are taught. Other personality traits, namely, Extraversion, Agreeableness, and Neuroticism failed to show any significant effects on learners' achievement and recall of structures.

With regard to the negative effects of TE, it could be claimed that, in the Iranian context, learners are mainly instructed using traditional materials, and the unfamiliarity of students and teachers with the TE materials moderates their effects. The curiosity and interest of students in TEGI, which provoked so many questions, shows the success of TE to attract students' attention. However, any claim about the depth and level of such attraction and attention is beyond the scope of this study. TE materials draw attention to specific features in the input; however, this awareness seems to be at the level of noticing rather than understanding, subsequently resulting in acquisition (Robinson, 1997; Vaez Dalili et al., 2011). Such attention may not necessarily be beneficial, but detrimental and distracting at some level due to the novelty of new materials in TE instruction or the multitude of other factors which can affect learners' attention (Park, 2004).

Moreover, it is clear the learners' IDs such as learning strategies and their benefits from different materials are not only affected by personal but also cultural and societal factors. These IDs can affect the meaning of each of the TE techniques and cues to learners. On the other hand, this study demonstrates that these TE cues attract attention, but whether this attention, in an overall meaningful communicative learning process, is attracting or distracting needs to be further investigated. To moderate such ID factors and enhance TE's beneficial effects on a larger educational scale, a systematic instruction of different TE cues could be advantageous. Triangulation of similar studies with think-aloud protocols, interviews, and possibly brain imaging systems could potentially provide detailed information about the quality and quantity of learners' attention and noticing while learning with TE materials.

Language learning is a slow process (Ellis, 1993), and the differences in various grammar teaching methods could be better investigated over longer periods of time (De Santis, 2008). The differences between the performances of learners may be due to different levels of noticing as a result of different task types (Leow, 200l) and demands (cf. Harley, 1994; Robinson, 2003; Schmidt, 1990; Simard & Wong, 2001). How do learners process different kinds of input either for meaning or form? When does their attentional capacity (VanPatten, 1996) become exhausted in a session of instruction with different formats of TE (cf. Balan, 1989; Hartley, 1982, 1986; Simard, 2009)? The effects of each of these TE materials on different aspects of language may be possible areas for future studies.

This study was done during the summer holidays in which learners' main instruction was through English classes designed for the purpose of the study. This, on the one hand, warrants the differences in the performances of the groups to be mainly the result of differences in teachers' instruction methods. On the other hand, learning is not limited to classroom instruction, but augmented by other sources of material outside of the classroom. Replication of this study over longer periods and during an educational year may lead to different results.

Finally, in this study a test was used as the assessment instrument. It is recommended that this test be substituted by elicitation tasks and more meaningful contextualized tools of language proficiency measurement in any future replications.

References

Abe, H. (2006). Input Enhancement in Pronunciation Pedagogy: The Impact on Learning Connected Speech in L2 English. *Research Reports of Tsuruoka National College of Technology, 41*, 15-38.

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and Awareness in Foreign Language Learning* (pp. 259-302). Honolulu: University of Hawaii Press.

Balan, P. (1989). Improving instructional print materials through text design. *Performance and Instruction, 28* (7), 13-18.

Bernard, R., & Scampton, D. (2008). Teaching Grammar: a Survey of EAP Teaching in New Zealand. *New Zealand studies in Applied Linguistics, 14*, 59-82

Costa, P. T., Jr., & McCrae, R. R. (1992a). *NEO PI-R professional manual*. Odessa, US FL: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1992b). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, US FL: Psychological Assessment Resources.

De Santis, P. (2008). Text enhancement and the acquisition of English verbal inflation –*s* by L1 Haitian Creole speakers. *Applied Language Learning*, *18* (1&2), 27-4

Digman, J. M. (1990). Personality structure: Emergence of the Five-Factor Model. *Annual Review of Psychology*, *41*, 417–440.

Doughty, C. (1991). Second language instruction does it make a difference. *Studies in second language acquisition*, *13*, 431-469

Doughty, C., & Williams, J. (1998). Issues and terminology. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp.1-12.). Cambridge, UK: Cambridge University Press.

Ellis, R. (2001). Investigating form-focused instruction. *Language Learning*, *51*. Supplement I 1-46.

Ellis, R. (1993). Second language acquisition and the structural syllabus. *TESOL Quarterly*, *27*, 91-113.

Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, *40* (1), 83–108.

Farahani, A. K., & Sarkhosh, M. (2012). Do Different Textual Enhancement Formats Have Differential Effects on the Intake of English Subjunctive Mood? *Theory and Practice in Language Studies*, *2* (4), 688-698. Retrieved from http://dx.doi.org/10.4304/tpls.2.4.688-698

Gass, S. M. (1997). *Input, interaction and the second language learner.* Mahwah, NJ, US: Lawrence Erlbaum.

Gass, S. M., & Selinker, L. (2008). *Second language acquisition.* New York, NY, US: Routledge.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicon. *Journal of Personality and Social Psychology*, *59*, 1216–1229.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*, 26–42.

Harley, B. (1994). Appealing to consciousness in the L2 classroom. *ALLA Review*, *11*, 57-68.

Hartley, J. (1982). Designing instructional text. In: Jonassen, D.H. (Ed.), *The Technology of Text*. Educational Technology publications (pp. 193–213), Englewood Cliffs, NJ, US.

Hartley, J. (1986). Planning the typographical structure of instructional text. *Educational Psychology*, *2*, 315–352.

Jin, J. (2011). An Evaluation of the Role of Consciousness in Second Language Learning. *International Journal of English Linguistics*, *1* (1).

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory—Versions 4a and 54*. Berkeley, US: University of California at Berkeley, Institute of Personality and Social Research.

John, O. P., & Naumann, L. P. (2007). *Correlations of BFI scales and self-reported act frequencies in an undergraduate sample*. Unpublished data, Institute of Personality and Social Research, University of California at Berkeley.

John, O. P., Robins, R., W., & Pervin, L., A. (2008). *Handbook of Personality* (3rd Ed.). The Guilford Press. New York.

John, O. P., & Soto, C. J. (2007). The importance of being valid: Reliability and the process of construct validation. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 461–494). New York, US: Guilford Press.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York, US: Guilford Press.

Jourdenais, R., Ota, M., Stauffer, S., Boyson, B., & Doughty, C. J. (1995). Does textual enhancement promote noticing? A think-aloud protocol analysis. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 183–216). Honolulu, US: University of Hawaii.

Kaufman, B., DeYoung, C.G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh N. (2010). Implicit learning as an ability. *Cognition*, *116*, 321-340.

Lee, S. K. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning*, *57*, 87-118.

Leeman, J., Arteogotia, I., Fridman, B., & Doughty, C. (1995). Integrating attention to form with meaning: Focus on Form in content based Spanish instruction. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 217-258). Honolulu, US: University of Hawaii Press.

Leow, R. P. (1997). The effects of input enhancement and text length on adult L2 readers' comprehension and intake in second language acquisition. *Applied Language Learning*, *8*(2), 151-182.

Leow, R. P. (2001). Do learners notice enhanced forms while interacting with the L2 input? An online and offline study of the role of written input enhancement inL2 reading. *Hispania, 84*, 496-509.

Leow, R., Egi, T., Nuevo, A., & Tsai, Y. (2003). The Roles of Textual Enhancement and Type of Linguistic Item in Adult L2 Learners' Comprehension and Intake. *Applied Language Learning, 13*(2), 1-16.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 413-468). San Diego, CA, US: Academic Press.

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research and practice. In C. J. Doughty, & J. Williams (Eds.). *Focus on form in classroom second language acquisition* (pp. 15-41). New York, US: Cambridge University Press.

McCrae, R. R., & Costa, P. T. (1990). *Personality in adulthood*. New York: Guilford Press.

Myers, I. (1962). *The Myers–Briggs Type Indicator*. Princeton, NJ, US: Educational Testing Service.

Overstreet, M. (1998). Text enhancement and content familiarity: The focus of learner attention. *Spanish Applied Linguistics, 2*, 229-258.

Pacheco, J. (2004). A comparison of textual enhancement and explicit rule presentation in secondary one English as a second language classes (unpublished master's thesis). Concordia University, Canada.

Park, E. S. (2004). Constraints of implicit focus on form: Insights from a study of input enhancement. *Teachers College, Columbia University Working Papers in TESOL and Applied Linguistics, 4*(2), 1-30.

Radwan, A. A. (2005). The effectiveness of explicit attention to form in language learning. *System, 33*(1), 69-87.

Rashtchi, M., & Gharanli, L. (2010). Noticing through Input Enhancement: Does it Affect Learning of the Conditionals? *Journal of Language and Translation, 1*(1)**,** 19-27.

Richards, J.C., Hull, J., & Proctor, S. (2005). *New Interchange 3* (3rd Ed.). Cambridge, UK: Cambridge University Press.

Robinson, P. (1995). Attention, memory, and the "Noticing" hypothesis. *Language Learning, 45*, 283–331.

Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning, 47*(1) 45-99.

Robinson, P. (2003). Attention and memory during SLA. In C. Doughty, & M. Long (Eds.), *Handbook of research in second language acquisition.* Oxford, UK: Blackwell.

Rosa, E., & O Neill, M. (1999). Explicitness, intake and the issue of awareness: Another piece to the puzzle. *Studies in Second Language Acquisition, 21* (4), 511-533.

Sarkhosh, M. & Sarboland, E. (2012). Different Textual Enhancement Formats and the Intake of English Past Simple Tense. *International Journal of Linguistics, 4* (2).

Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment*, *63*, 506–516.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied linguistics, 11,* 129-158. doi:*10.1093/applin/11.2.129*

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics, 13,* 206–26.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. In: Hulstijn, J., Schmidt, R. (Eds.), *Consciousness and Second Language Learning: Conceptual, Methodological and Practical Issues in Language Learning and Teaching, 11, AILA Review*, pp. 11–26.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in language learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu, HI: University of Hawaii Press.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge, UK: Cambridge University Press.

Sharwood Smith, M. (1981). Consciousness raising and the second language learner. *Applied Linguistics, 2*, 159-168. http://dx.doi.org/10.1093/applin/2.2.159

Sharwood Smith, M. (1991). Speaking to many minds: on the relevance of different types of language information for the L2 learner. *Second Language Research, 7,* 118-32.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition, 15* (2), 156-179.

Shook, D. J. (1994). FL/L2 reading, grammatical information, and the input-to-intake phenomenon. *Applied Language Learning, 5*, 57–93.

Simard, *D.*, & Wong, W. (2001). Alertness, orientation, and detection. *Studies in Second Language Acquisition, 23*, 103-124.

Simard, D. (2009). Differential effects of textual enhancement formats on intake. *System, 37*(1), 124–135.

Tomlin, R.S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 16*, 183–203.

Vaez Dalili, M., Ketabi, S., Kassaian, Z., & Eslami Rasekh, A. (2011). The Effect of Textual Enhancement on System Learning of English Dative Alternation. *World Applied Sciences Journal, 15,* 63-69

VanPatten, B. (1996). *Input Processing and Grammar Instruction: Theory and Research.* Norwood, NJ, US: Ablex.

Wong, W. Y. (2000). The effects of textual enhancement and simplified input on L2 comprehension and acquisition of non-meaningful grammatical form. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, Champaign, IL.

Appendix

A sample of TE materials used in the study.
- **What would you do if you found 10 million dollars**?
-Would you return it to its owner?
-Would you be happy if you lost all your money?

Read the following conversation.
Phil: Look at this. Some guy found $750,000! He returned it and the owner simply thanked him with a phone call.
Pat: You are kidding**. If I found $750,000, I wouldn't return it so fast**.
Phil: Why? What would you do?
Pat: Well, I would go straight to the mall and spend it. I could buy lots of nice clothes and jewelry.
Phil: Someone might also find out about it. And then you could go to jail.
Pat: Hmm. You've got a point there.
Pat: **What would you do if you found that much money**?
Phil: If I found $750,000 dollars, I would return it.
Pat**: What would you do if you couldn't find the owner**?
Phil: **If I couldn't find the owner, I would give it to charity**.
Pat: Then you must be a good person, but **if I were the owner of that money, I would give $50,000 to the finder.**
Phil: Yeah. I think people who return lost things should get a reward.

Q1. What is the best thing to do with lost things?
Q2. What would Phil do if they found lots of money?

Using Pre-Task Models to Promote Mining in Task-Based Language Teaching

Martin Hawkes
The University of Shiga Prefecture, Japan

Abstract

Practitioners of task-based language teaching often use pre-task activities to prime learners before they perform meaning-focused tasks. Such activities can be used to introduce topics, present useful vocabulary, or, more controversially, target specific forms. It has been claimed that learners will self-select, or *mine*, useful language during this stage. This investigation looked at how learners used one kind of pre-task activity—task models performed by proficient speakers—to help them prepare for the main task performance. The study took place at a Japanese university with three sections of first-year high beginner learners. Two task types were used. Section A listened to a task model, answered comprehension questions, then performed a listening cloze exercise with a transcript of the conversation. Section B only listened to a task model and answered comprehension questions. Section C performed tasks with no task model. The task interaction was recorded, transcribed, and analysed qualitatively. Results showed there was little mining from the participants who only listened to the models, but there was evidence of mining from those who performed the cloze. There was also some indication that pre-task modeling can positively affect strategy and patterns of interaction. The findings indicated that pre-task modelling may be a useful option for teachers using tasks in their classrooms.

Keywords: TBLT, pre-task, models, mining

Introduction

The use of tasks in language learning and teaching has remained a popular topic for both SLA research and pedagogy. As well as being featured prominently in articles in teaching English to speakers of other languages (TESOL) and applied linguistics journals, task-based language teaching (TBLT) has also been the subject of both pedagogically-focused (Edwards & Willis, 2005; Nunan, 2004; Willis, 1996; Willis & Willis, 2007) and research-focused (Ellis, 2003; Robinson, 2011; Samuda & Bygate, 2008) book-length treatments, international conferences, and smaller specialist publications.

Two issues that have received much attention are how tasks should be sequenced within a set of classroom activities, and how form-focused instruction ought to be integrated with TBLT. Most proponents of TBLT have argued that pre-task activities are essential to prime learners for the upcoming main task. One purpose of these pre-task activities is to provide input that learners *mine* for potentially useful language.

This paper investigates the topic of mining by lower-level Japanese university students. One kind of pre-task activity is for learners to listen to a model of the task they are about to perform themselves. This study examines the extent that learners mine these task models for helpful language and use it in their own task interaction.

* *Tel. +81(0)749-28-8253; Email: hawkes.m@shc.usp.ac.jp; Address: School of Human Cultures, The University of Shiga Prefecture, 2500 Hassakacho, Hikone City, Shiga, Japan*

Tasks in language teaching

Originally, TBLT grew organically as an offshoot of communicative language teaching. This was perhaps due to disaffection among some practitioners and researchers with what was seen as pseudo-communicative approaches to language teaching exemplified most famously by the presentation-practice-production (P-P-P) classroom sequencing. In a typical P-P-P procedure, language classes follow a grammar-based syllabus, and each individual lesson features a *form-of-the-day*, which learners are expected to master by the end. First, the teacher presents an isolated target form with exemplars. Next, learners typically perform some controlled practice of the target language. Finally, the learners carry out a communicative activity, which is designed to necessitate the target form, and they are expected to demonstrate some degree of proficiency with it. P-P-P remains a popular methodological choice for many practitioners in the East Asia region (Carless, 2009; Long & Kurzwell, 2002; Sato, 2010). Moreover, advocates of P-P-P have a psycho-linguistic rationale for the approach in skill acquisition theory (Anderson, 2010; DeKeyser, 1998). Skill acquisition theory argues that language is no different from other human skills such as driving a car or playing a musical instrument. First, learners gain *declarative* knowledge of a linguistic form, which is then *proceduralised* through intensive practice. This eventually leads to *automatisation*, when learners are able to accurately produce the language point with a high degree of fluency.

However, traditional communicative approaches like P-P-P have come under attack for several years now by SLA researchers. Vocal critics like Skehan (1996) have argued that any approach that has a fixed grammar-oriented syllabus, and expects mastery of these target forms, is doomed to fail because it is incompatible with what is known about developmental sequences. Quite simply, second language learners cannot learn grammatical forms for which they are not ready; instruction has been shown to have a negligible effect, at best, on the order that forms are acquired (Klapper, 2003; Ortega, 2011). Further, practitioner-oriented critics of P-P-P have attacked the way that classroom procedures are sequenced. Willis and Willis (2007) claimed that if learners are instructed to use a specified target form during a communicative activity, there will not be a focus on meaning; instead, learners will merely reproduce the target forms for display purposes. Conversely, they may simply ignore the target forms, rendering the presentation and practice stages meaningless (Larsen-Freeman, 2009; Thornbury, 1997). TBLT is arguably the approach that has been the most attractive alternative for many researchers and practitioners.

In TBLT, a central tenet is that lessons (and also syllabuses in stronger versions) should be based around *tasks*. But, what exactly are tasks, and how can they be defined? Over the years, there have been several attempts to define the essence of a task (see for example Breen, 1987; Ellis, 2003, pp. 2-10; Long, 1985; Nunan, 2007; Samuda & Bygate, 2008, pp. 62-70). For the purpose of this paper, tasks are defined following Skehan (1998), who suggested that they must possess the following five characteristics:

1. Meaning is primary
2. Learners are not given other people's meaning to regurgitate
3. There is some sort of relationship to comparable real-world activities
4. Task completion has some priority
5. The assessment of the task is in terms of outcome. (p. 95)

While Skehan's definition can be applied to all four skills, most research has looked at oral communicative tasks. But, TBLT is not simply the use of these meaning-focused activities. As the discussion of the P-P-P procedure mentioned above, how tasks are used and sequenced has a great effect on how learners approach them.

Task Frameworks and the Importance of Pre-Task Activities

One point that may be useful to emphasise is the need for some kind of focus on language. Although there may be some that argue for a purely meaning-focused approach to TBLT that eschews any focus on language—a zero-grammar approach (Ellis, 2005)—most researchers seem to be in agreement that some kind of attention to form is necessary for interlanguage development (Burrows, 2008; Doughty & Williams, 1998), and to prevent fossilisation and pidginisation (Johnson, 1996; Willis & Willis, 2007). As a result, several frameworks for using tasks in language teaching have been proposed, all of which have a focus on language form built into them.

Sequences of Tasks with a Post-Task Explicit Language Focus

The 2007 approach proposed by Willis and Willis developed the earlier framework of J. Willis (1996). Willis and Willis (2007) stated that a task should rarely exist in isolation, but it should be integrated into a sequence of tasks and activities. This sequence begins with a *priming* or *pre-task* stage, which is flexible and can take many forms. The purpose of this stage is to introduce the topic of the *main task(s)* and to highlight potentially useful language. However, there is no requirement for formal accuracy at this stage. Some examples they gave of priming included the following: teacher-led discussions, listening to a recording related to the topic, memory games to highlight relevant vocabulary, questionnaires, brainstorming, writing potentially useful vocabulary and/or phrases on the board, and listening to a model of the main task that the learners will later do themselves. Following the priming stage, there is one or more main task, which is usually followed by planning and a report by some students to the class about the results of the task. Finally, Willis and Willis argued that some kind of explicit focus on specific language forms is necessary, but only at the end of the task sequence.

Sequences of Tasks of Increasing Complexity with Reactive Focus-on-Form

Long's (2015) approach to TBLT is similar to Willis and Willis' (2007) in many ways. Long (2015) also argued for sequences of pedagogic tasks with a focus on meaning, following Robinson's (2005) proposal of sequencing tasks in order of complexity. The first task(s), similar to Willis and Willis' approach, should be for priming. Although these tasks should not feature isolated grammatical forms to be explicitly pre-taught, Long encouraged teachers to draw learners' attention towards lexical items and collocations. This can be achieved through *input enhancement* (Sharwood Smith, 1993). One means Long (2015) suggested is to highlight useful language in written texts using bold or italicised typeface. The remainder of the pedagogic tasks gradually increase in complexity, and although there is no place for pre-planned explicit teaching of forms in Long's framework, he did advocate a *focus-on-form* (Long, 1991), in which language points are tackled reactively as problems arise.

Sequences of Activities with a Pre-Task Explicit Language Focus

Nunan's (2004) approach to TBLT is fundamentally different to both Long (2015), and Willis and Willis (2007). The main difference lies in the order of the 'task' sequence. Nunan's framework begins with a *schema building* stage, in which useful vocabulary and expressions are highlighted. Next, learners listen to and practice sample conversations before the explicit teaching of some target forms. It is only after this stage that the learners partake in something most practitioners would recognise as a communicative task.

There are, no doubt, other suggestions of how to incorporate tasks in language teaching, but what these three approaches demonstrate is the differences in interpretation within the field of TBLT. One aspect that seems to be agreed upon, however, is the need for pre-task activities to help learners prepare for their upcoming task performance.

Skehan (1996) suggested that pre-task activities could assist learners by either reducing the cognitive or linguistic demands of the main task. By lowering cognitive demands, learners are able to attend more to language form during the main task. Skehan identified the following four broad approaches to achieve this: 1) *foregrounding,* where a teacher may introduce a topic through a class discussion or by telling a story; 2) *doing similar tasks,* which often leads to greater fluency, complexity, and richer vocabulary use (for a summary of research into task repetition, see Ahmadian, 2012); 3) *planning,* that is, providing learners with planning time to lead to greater fluency, complexity, and to a lesser extent accuracy (Foster & Skehan, 1996; Skehan & Foster, 2005); and 4) *observing,* that is, having learners listen to or watch two proficient speakers of the target language performing the same task they will subsequently do themselves. This has the goal of allowing learners to attend to form and reducing the need to think about task procedures. Finally, Skehan (1996) proposed that a reduction in the linguistic demands of a task could be achieved by explicit or implicit teaching, or through consciousness raising.

Ellis (2003) suggested that the observation of a model task could be supplemented by consciousness raising of particular linguistic features present in the task model interaction.

Essentially, the pre-task activities described above aim to implicitly direct learners' attention to linguistic features of some kind. When learners are focused solely on communication, they are unlikely to notice certain language points that may be useful for the upcoming task. According to Schmidt's noticing hypothesis (1990, 2001), the noticing of linguistic forms from input—which requires attention—is necessary for acquisition. However, if the attention to forms is too explicit, it may have a detrimental effect on the communicative nature of a task. Indeed, both Skehan (1996) and Ellis (2003) cautioned that any focus on specific language forms could potentially damage the integrity of the task, leading learners to no longer focus primarily on meaning, and instead treat the task as simply an exercise to practice the highlighted target forms. What is needed is an implicit focus on useful language in the pre-task stage. Ellis (2003), Skehan (1996), and Willis and Willis (2007) all suggested that pre-task work with task models may be useful for this purpose.

Mining

Mining is the term used when learners identify and select useful language from pre-task materials and incorporate it into their task performance. Willis and Willis (2007) claimed that this is very different from a teacher dictating a form-of-the-day and requiring a learner to use it during a classroom activity; thus, in their view, Skehan (1996) and Ellis's (2003) fear of tasks losing their focus on meaning is not applicable when it comes to mined language self-selected by learners. Recently, there have been some studies looking at mining from pre-task, and on-task, materials. Boston (2008), in a context with lower level Japanese university students, found that while learners did not mine language from pre-task listening models, they did so from written task instructions. Boston concluded that it may be possible to plant target forms into task materials in order to provide a form focus. In a Korean junior high school context, Kim and McDonough (2011) found that although pre-task modelling led to increased mining from task materials, learners were unlikely to copy language verbatim, picking out only key words or expressions instead. Kim (2013), in the same context, reported that learners who watched a model task with an emphasis on question forms produced more target-like questions during the main task, demonstrating that pre-task modelling may be useful to direct learners' attention towards specific forms.

This paper aims to further investigate these topics; specifically, it looks at mining from listening to pre-task models and language that has been highlighted in written pre-task materials. With these issues in mind, the following research questions were formulated for this study:

1. Did learners mine language after listening to task models and processing them for meaning?
2. Did learners mine language after performing an additional listening cloze exercise using task model transcripts?

Methods

Participants

The context of this study was a compulsory 15-week basic English communication course for first year, non-English majors at a Japanese university. Each class was 90 minutes, and the stated purpose was to improve students' speaking skills. There were 58 participants that agreed to take part in the study: 32 females and 26 males. Participants had fairly limited fluency and implicit grammatical knowledge although, as is common in Japanese university contexts, they had a relatively high explicit grammatical knowledge. Despite this, they would probably be classed as high beginner learners, similar to the learners in Boston's (2008) study. All first year students had taken an in-house placement test to divide them into three bands, and then they had been randomly assigned to classes within each band. Unfortunately, the test data was confidential and not released for the purposes of this study. The participating classes were all from the middle band. In the course, students completed six units from the compulsory textbook *New Interchange 1A*. A unit took two weeks to complete, but each was supplemented by communicative tasks related to the current topics. The author was the teacher of these classes and had five years of experience teaching the same course in the same context.

Tasks

The previous year, a pilot study examining the task interaction by students in the same context was performed with a variety of tasks. Two of these tasks were chosen for the study, which, following the taxonomy devised by Pica, Kanagy, and Falodun (1993), would be classed as a decision making task and a jigsaw task. Task 1, a decision making task, involved groups of three or four students planning a trip to the cinema. Using a copy of the current week's cinema schedule, they had to decide which film and showing to see together. Next, they needed to choose a good place to eat, and whether they should dine before or after the movie. They also had to use their knowledge of the area to organise when and where to meet. There was no time limit for students to finish, but the researcher had often used this task in the same teaching context. From the experience of observing similar learners performing this task many times, the researcher had established that the task usually takes four to six minutes to successfully complete. Task 2 was the jigsaw task in which, participants in groups of four were each given a schedule for the next seven days. In the task scenario, they had to find a mutually convenient time to meet and do a homework assignment. This task was similar to scheduling tasks commonly found in published materials (e.g., Barnard, Cady, Duckworth, & Trew, 2009).

Task Models

For each task, a model was recorded by pairs of proficient English speakers who were colleagues of the researcher. To enhance authenticity, no script or guidance regarding what language forms should be used was given. The task materials were provided, and the speakers were asked to perform the task as naturally as possible. The task interaction was audio recorded and transcribed orthographically for ease of comprehension for the participants. The researcher identified some key expressions and deleted them for use in the listening cloze activity. The transcripts of the model tasks can be found in the Appendix.

Procedures

Three classes were chosen to participate, and each was assigned a different treatment. The classes consisted of 16 to 19 participants, who were divided into groups of three or four to perform the tasks, depending on attendance on the day. Table 1 shows the number of groups for each section and task. Section A listened twice to an audio recording of a task model, and answered some comprehension questions. Following this, they received a transcript with the key expressions deleted. The model was played twice again (four times in total), and participants completed a listening cloze exercise. This was a form of input enhancement designed to increase the chances of noticing by implicitly guiding participants' attention to specific expressions. At no point were participants instructed to use these forms in the main task. Section B completed the first part of these activities, that is, they listened to the model and answered comprehension questions only. Section C neither listened to the model, nor saw the transcript before the task, although they did have the opportunity to do these activities *after* the main task. This stage was followed by the main communicative task. The participants in Sections A and B were instructed not to use the language from the model. Therefore, any language that was produced in the main task would be self-selected and could be considered as evidence of mining. Audio recordings were made of the main task interaction. This was then transcribed in detail and analysed qualitatively to look for any effects of task models on the content of task interaction.

Evidence of mining was primarily operationalised by the presence of full or partial use of the key expressions found in the task models. Analysis of the task performances in the pilot study suggested that these key expressions would be useful for task completion. The pilot data also indicated that the expressions were unlikely to be within the participants' productive knowledge; previous students in the same context had consistently avoided or made errors with the functions that the key expressions realised. Therefore, their use might be evidence of mining from the pre-task activities. The remainder of the task interaction was scrutinised carefully to look for examples of other single or multi-word lexical items—often accompanied by unfilled pauses and other disfluencies that suggested word searching—that corresponded to the task models.

Table 1
Summarized Description of Each Section

	Section A	Section B	Section C
Pre-task activities	Listening (twice) + comprehension questions; listening (twice) + cloze exercise	Listening (twice) + comprehension questions	No task modelling
Number of groups in Task 1	5	4	5
Number of groups in Task 2	4	5	4

Results

One characteristic of the data that quickly became apparent for both tasks was the lack of mining by participants in Section B and Section A of language other than the key expressions that were deleted in the transcripts. Therefore, the following discussion will largely focus on the language used in the cloze exercises and the extent to which it was mined by the participants in Section A. I will also describe whether participants in the other sections used these expressions and, if not, what alternatives they used and with what level of success. To protect the anonymity of the participants, pseudonyms have been used in this section.

Task 1: Cinema trip (decision making)

The opening line of the task model saw Rod orienting himself and his partner, Jim, to the task with the following utterance (underlined words were deleted in the cloze exercise):

Rod: So Jim, <u>do you want to see a movie</u> this weekend?

In four of the five groups in Section A, their opening utterance appeared to involve some mining from the model. In one group, Ryu used the deleted words verbatim, as shown in Excerpt 1.

```
(1)   R: do you want to see a movie?
      K: I want to watch thee- the Lesson of the Evil
```

In Excerpt 2, Ichiro not only mined the deleted words but also the "this weekend" that followed in the model. Ichiro then further specified the day, perhaps remembering the task instructions, which stated that the group were to meet on the upcoming Saturday.

```
(2)   I: do you want to see a movie this week? weekend? Sat-
         on Saturday?
      T: ah::: I want to see er::: Evangelion.
```

Excerpt 3 shows how, in a third group, Miyako and Takahiro made an unsuccessful attempt to begin the task that ended in laughter then a period of silence. Kazuki broke this silence with a fluent effort to reproduce the model's opening, even placing "everyone" into the name position in the model.

```
(3)  M: what [movie
     T:      [what movie
     ((laughter))
     (8.0)
     K: so everyone do you want to see ((laughs)) a movie this
        weekend?
```

The members of a fourth group did not begin their performance by mining this question. However, Excerpt 4 shows how Teppei did not seem to accept Kouta's opening utterance as satisfactory by adding "this weekend" to the beginning of the response, making their conversation conform more closely to the model's opening.

```
(4)  K: what do you wanna see (..) movie?
     T: this weekend? um (2.0) I want to: (.) I want to see a movie
        (.) Les Miserables
```

No participants from Sections B and C used the same expression found in the model. All groups opened with a relatively successful attempt of "what movie do you want to see?", or by a participant stating their preference, as in the following example from Section B:

```
(5)  S: I want to see Chugakusei Maruyama
     Y: I want to huh? (..) I want to see Library Wars
```

This suggests that while the model arguably reflects how a natural discussion may begin, the participants were very much oriented towards task completion, for which the first goal was to choose a film to watch.

This decision making task created several obligatory occasions where participants had to make suggestions. In the model, the speakers used "let's", "what about?", and "how about?" to do this. For example, Jim made the following suggestion, which was used as part of the cloze:

Jim: Brilliant okay, so <u>how about</u> the er two-twenty?

However, there was only one occasion when a participant in Section A used "how about?" to make a suggestion (Excerpt 6), making it difficult to ascertain whether it was due to mining of the model.

```
(6)  T: okay how about Les Miserables? miru miru
        ((translation: watch it?))
```

Again, no participants in Sections B and C used this expression in their task performances. Throughout the data, there were only six examples of participants using "let's", an appropriate expression for making suggestions. Participants overwhelmingly made pragmatically inappropriate suggestions by saying "I want…" or by making bald statements which were sometimes, but by no means always, accompanied by rising intonation. This is illustrated in the last line of excerpt 7, where Dai suggested meeting at the university. Using appropriate expressions for making suggestions would certainly have helped with the quality of their task interaction.

```
(7)  C: okay first er:: (..) we (.) where we meet?
     (3.5)
```

```
     K: meet
     C: meet
     K: meet (1.5) eh?
     D: university?
```

One item that clearly looked to have been mined was the following question form found in the model:

Rod: But <u>where should we meet?</u>

Three of the groups in Section A used it in their task performance. Excerpt 8 shows that, following a long period of whispering and silence, Ken finally reproduced the deleted words from the model. It is likely that Ken consulted the model transcript during the period of silence to find this question, which presumably they were having some difficulty determining how to formulate.

```
(8)  A: we (.) go to (..) restaurant (.) for example (.)
        Gusto
     (12.0)
     ((unidentified unintelligible whispering in L1))
     R: okay
     A: okay
     (7.0)
     K: where should we meet?
```

Excerpt 9 also shows a period of whispering before this question was asked by Eri. Kazuki appeared to have misunderstood Eri's original question, and the whispered exchange in L1 seemed to have repaired the issue, leading Eri to continue the task with the question from the model.

```
(9)  E: when meet wa? (tr: wa is a Japanese topic marker)
     K: eat (.) after movie
     ((unidentified unintelligible whispering in L1))
     E: but (.) but (.) where should we meet?
```

Finally, in Excerpt 10, Teppei used the question form accurately after a short period of silence and a hesitation device. It is possible he was consulting the model transcript during this time.

```
(10) T: Sannomiya station (1.5) uh::: where should we meet? (1.0)
        uh: at Sannomiya station?
     K: yeah
```

No examples could be found in the other section groups' task interaction of this question form. In fact, when participants asked about the meeting place, in all the data set, there were no accurately produced questions. Most attempts either featured minimalisation of the like illustrated in excerpts 7 and 9. In Section C, two of the four groups even resorted to using the L1 to raise the topic, as illustrated in excerpt 11.

```
(11) A: machiawase basho ((tr: meeting place))
     C: machiawase basho ah okay=
     A: =um Hankyu Umeda
```

Furthermore, although this question form was clearly mined by a member in three of the groups in section A when they were asking about the meeting place, none of these participants applied it elsewhere, for example, to ask about the meeting time. This indicates that whereas some participants appeared to mine this language chunk, they did not take the question frame and apply it when they had other suitable opportunities.

Other expressions for which there was some evidence of mining were "that sounds great" and perhaps "sounds pretty cool", which may have been useful for participants to support or agree to a suggestion.

```
Rod: Ah oh 400 yen but that's okay ah yeah 3D sounds pretty cool.
```

```
Rod: Ah that sounds great. Do you know any places around there?
```

In Group 3, two participants used "sounds good" consecutively when discussing the type of food to eat (excerpt 12).

```
(12) K: hm I want to eat (..) hm (2.5) Chinese
     R: Chinese (..) [sounds good.
     T:              [ahh
     (1.5)
     T: sounds good (.) Chinese it sounds good
```

Once more, there were no examples of participants from Sections B and C using something close to these expressions for agreeing. In the vast majority of cases where such expressions would have been suitable, participants tended to use "yeah", "okay", and "me too" (for example, see Excerpt 13). While these responses could sometimes sound appropriate, expressions of the type used in the model could have enriched the participants' language options.

```
(13) T: let's see Zekkyo-Gakkyu
     H: yeah okay
```

The final deleted expression of some interest was the following rejection by Jim of a suggestion by Rod:

```
Jim: Ooh I'm not really into animation.
```

This expression was not mined by any participants in Sections A or B, and was not used by any participant in the third section either. Instead, participants used a variety of expressions to disagree or reject suggestions including "I don't like...", "I'm not interested in...", "I want to...", and simply "no". While some of these sounded perfectly appropriate, sometimes deploying an indirect rejection of an idea like the one used by the native speakers in the model could be extremely useful as a face-saving measure for one's interlocutor. An interesting point that was revealed through examining the language used for agreeing with or rejecting suggestions was the paucity of such exchanges in the groups of Section C. All four groups followed the first movie suggestion that was made and promptly moved onto the next topic to discuss. Indeed, in all the data for Section C, there was only one example where a participant rejected a suggestion of *any* kind (Excerpt 14).

```
(14) A: lunch
     T: hm::: (2.0) McDonald
     A: Chieko
     C: he::: McDonald? I want to eat (.) pizza!
```

This suggests that one impact of pre-task modelling for this task was to encourage participants to discuss the different options available and to challenge others' suggestions. It is possible that although the participants in

Section B did not mine any language form, there is some evidence that it brought the participants' task strategy closer to the intentions of the teacher.

Although it is clear that some participants incorporated items from the cloze into their task performances, it certainly cannot be claimed that all or even most participants did. Of the 14 participants in the four groups in Section A, only seven used the expressions from the cloze, and of these seven, not all used the forms on every possible occasion. They mostly elected to use language, often incorrectly, from their own resources. Examples of other language mined from the transcripts were even more scarce. In fact, in all the Section A data, there was only one example that was considered likely to be mined—the use of "lobby" as the meeting place (excerpt 15), which was the same meeting place decided on in the model.

```
(15) T: where (.) do we meet?
     H: er (..) lobby?
     T: lobby?
     Y: lobby?
     H: movie er (..) theatre mobby? theatre lobby?
     T: okay.
     H: okay?
     T: okay I understand
```

Task 2: Scheduling task (jigsaw)

The deleted items in the cloze exercise for Task 2 can be broken in two categories: making suggestions and stating plans. In the model, performed by Mark and Olive, Mark used the question frame "how is...for you?" to make the following suggestion:

Mark: So we need to organise a time to meet, em, so <u>how is Thursday morning for you</u>?

Of the four groups in Section A, only one mined this question frame and applied it throughout their task performance. In excerpt 16, it can be seen how Yoshi, after spending some time trying to recall an appropriate way to make a suggestion, seemingly consulted the transcript and used the whole frame to begin the task.

```
(16) Y: how do you hm? how: do: you: how- how is Sunday morning for
        you?
```

In the remainder of the task, all four members of the group used "how is...?" (without "for you") on sixteen occasions to make suggestions, illustrating how learners can also mine language from each other during task interaction.

As in Task 1, the expression "how about...?" featured in the following suggestion made in the task model:

Olive: Ah <u>how about Monday morning</u>?

However, unlike in the cinema task, where this same expression for making suggestions was used by only a single participant, here it was used frequently, 13 times by six different participants from the three other groups. There was some other use of this chunk from outside of Section A: In Section B, where it was used on one occasion by two participants, and in section C, where it was used four times by a single participant. Outside of Section A, the attempts at suggestions were mostly similar to the bald suggestions and minimalisation that was present in the Task 1 data. In addition, some groups chose to follow a different strategy towards task completion than the two speakers in the model. Instead of suggesting days and times to meet, participants took turns stating all the times they were available (excerpt 17). This strategy was followed by one group in Section A, and two groups in Section

B. When this strategy was employed, there was no need to either make suggestions or talk about the plans or arrangements that they already had.

```
(17) M: I have (.) free time (.) on (.) Sunday morning er (1.5)
        Wednesday morning (1.0) Thursday evening (.) er and Friday
        morning.
     H: I have free time is (..) Sunday afternoon (1.0) Monday
        evening (..) Wednesday morning (.) and Friday morning
     N: er I have free time on (.) Wednesday (.) and (.) Friday:
        morning (..) and: Monday evening (.) and Wednesday evening
```

Moreover, two groups in section C avoided making suggestions by using more direct questions to ask whether their group members were free (excerpt 18). However, unlike the previous alternative strategy, these groups still gave reasons for rejecting suggestions of date and time.

```
(18) H: I er do you (..) free (.) Monday evening?
     J: no I have PT job
     H: do you free (.) Tuesday afternoon?
     Y: no (.) I- I going to (.) dentist
```

When rejecting suggestions, the speakers in the model always gave a reason why a particular time was not suitable. For example, Olive used "have to" to show that she had firm plans for the Thursday evening:

Olive: Actually, I can't do Thursday evening <u>I have to cook dinner</u>.

She also used "gotta" in the same way:

Olive: Sunday night won't work at all. I have some family meeting
that I <u>gotta</u> go to.

The only participants from section A to mine this were three from a group who used "have to" on eight occasions. It was sometimes used accurately, or, as illustrated by excerpt 19, sometimes with the lexical verb omitted. There was one example of "have to" being used by a participant in Section B, making it unclear whether this form had indeed been mined. All other efforts to state plans would be deemed non-target-like errors with frequent minimalisation.

```
(19) K: how is Wednesday evening?
     D: Wednesday evening?
     Y: I'm PT job- I have to PT job.
     I: I have to write important essay
```

Overall, for Task 2, there was evidence that participants mined some of the language present in the cloze and deployed it during their task performance. This was mostly done by the four members of one group, while another group did not use any. Of the 15 participants in Section A, nine used language found in the cloze. Unlike Task 1, in Task 2 the groups of Section B and C used some of the cloze language, albeit not with the same frequency or range as by Section A participants. Of the fourteen participants in Section B, four used some of the cloze language. Further, of the sixteen participants in Section C, there were another four users of the cloze items. Despite being used much less than by the Section A participants, the very presence of these items in the other sections casts some doubt as to whether they were indeed mined.

In the Section-A group that did not mine any cloze expressions, one participant took the following opening line of the model transcript and used it to begin their task.

```
Mark: Okay so we need to finish this project by next Tuesday.
```

However, when they used it, at least one participant seemed to not understand the meaning, and the model was abandoned as a resource for the remainder of the task performance (Excerpt 20). This was the only example of any kind of mining by this group, and it was a rare example of a group mining language other than that found in the cloze.

```
(20) Y: we need to finish this project.
     (2.0)
     A: huh?
     (7.0)
     A: my- my free time (.) is (.) Monday evening
```

Table 2 shows a summary of the frequency with which key expressions were mined by the participants in this study. Section A consistently used these items over the two tasks. The higher number of uses for the task B items was due to the greater number of obligatory occasions that this task inherently presented. There is a clear difference between the frequency with which the key expressions were used by the participants in section A, who did the extra cloze exercise, compared with section B, who did not. In fact, there seems to be no discernable differences between section B and section C, who did not even listen to the task model.

Table 2
Frequency of Uses of Key Expressions During Task Performances

	Section A	Section B	Section C
Task A			
do you want to see a movie?	4	0	0
how about…	1	0	0
where should we meet?	3	0	0
sounds…	2	0	0
I'm not really into…	0	0	0
Task B			
how is…	16	0	0
how about…	13	2	4
I have to…	8	1	0
I gotta…	0	0	0

Discussion and Conclusion

The first research question asked whether learners mined language after listening to task models. The data collected from Section B and Section C showed that participants who had only listened to the model (without seeing the transcript or performing a cloze exercise) did not seem to mine language from the model more than those who did not listen to any model at all. This follows the findings of Boston (2008) that lower level learners do not seem to mine language from previous exposure to listening texts alone.

The second research question looked at the extent of mining that might occur after learners had done a listening cloze exercise using the transcripts of the model task. The data from Section A and Section B indicate that those who performed the listening cloze did mine language from the deleted expressions that were focused on in the listening activity. However, the possibility also exists that the observed mining was due to the two extra opportunities that section A had to listen to the task model. It is also possible that some may have used these language forms anyway as some of the cloze language was used by participants in the other sections that did not perform the cloze exercise, see the transcript, or even listen to the task model. Furthermore, mining was not particularly widespread, and many alternative language forms were used alongside those that did appear to be mined. Even when participants had access to written transcripts of the task model during their own task performances, they scarcely incorporated other expressions that would have been an improvement over their own resources. Therefore, there is little support here for the idea that language learners might treat the task as merely an exercise to practice certain target forms, suggesting that the use of pre-task modelling is consistent with one of the fundamentals of TBLT; that is, the primary focus of the task should be on meaning. Only occasionally did participants use language that they appeared to have mined from the pre-task model.

The participants in Section C did not see the model in any form and so performed the task entirely using their own (collective) means. Although their language resources were sufficient for successful task completion, there was evidence of an effect on how these groups approached the task. In Task 1, the decision making task, the groups in Section C did not discuss the options to the same extent as the participants of other sections that had listened to the model. This is possibly a sign of a focus on task completion that Seedhouse (2004) described as being a typical feature of task-based interaction: when learners are focused towards task completion, they tend to take the most efficient route to reach this goal. This, along with the alternative strategies taken by some groups in Task 2, highlights the conflict between task-as-workplan and task-as-process (Breen, 1989); the propensity for learners to interpret and perform a task differently to how it was intended by their teacher has been described in TBLT research (Kumaravadivelu, 1991).

Overall, the vast majority of forms employed by the participants were non-target-like. One notable feature was of language minimilisation. There were copious examples of participants omitting words, especially first person subjects, copulas, and auxiliary verbs. This was identified by Seedhouse (1999) as another characteristic of task-based interaction, as he described how there is a "general tendency to minimize the volume of language used, and to produce only that which is necessary to complete the task" (p. 152).

The findings of this study tentatively suggest that task models, supported by activities using transcripts, could be useful for directing learners' attention towards specific forms. However, there are a number of limitations to this study. First, in some cases, it is difficult to conclusively determine whether a participant mined language from the materials, or whether it already existed in their language repertoire. To investigate this further, some kind of introspective data collection tool such as stimulated recall protocols could be used to ascertain from where the participants retrieved the language forms or expressions. The use of video, instead of the audio-only recordings in this investigation, might also be illuminating in cases where it was possible that participants in Section A had consulted the model transcripts. Next, the number of participants in the study was relatively small, so it is not easy to know whether those few participants from Section B and C who used cloze expressions were exceptional, or whether many learners in this context would be comfortable with their use. Finally, it is clear that the task models did not have a significant impact on the ability of groups to successfully complete the tasks. The two tasks used in this investigation seemed to be fairly straightforward for these learners in terms of difficulty and complexity. It would be interesting to see whether, when given a more challenging task, learners would feel the need to mine language or pay more attention to the strategy used in a pre-task model.

References

Ahmadian, M. J. (2012). Task repetition in ELT. *ELT Journal, 66*(3), 380-382.

Anderson, J. R. (2010). *Cognitive psychology and its implications* (7th ed.). New York, NY: Worth Publishers.

Barnard, R., Cady, J., Duckworth, M., & Trew, G. (2009). *Business venture 2* (3rd ed.). Oxford, UK: Oxford University Press.

Boston, J. S. (2008). Learner mining of pre-task and task input. *ELT Journal, 62*(1), 66-76.

Breen, M. P. (1987). Learner contributions to task design. In C. Candlin and D. Murphy (Eds.), *Language learning tasks* (pp. 23-46). Englewood Cliffs, NJ: Prentice-Hall.

Breen, M. (1989). The evaluation cycle for language learning tasks. In R. K. Johnson (Ed.), *The second language curriculum* (pp. 187-206). Cambridge, UK: Cambridge University Press.

Burrows, C. (2008). Socio-cultural barriers facing TBL in Japan. *The Language Teacher 32*(8), 15-19.

Carless, D. (2009). Revisiting the TBLT versus P-P-P debate: Voices from Hong Kong. *Asian Journal of English Language Teaching 19*, 49-66.

DeKeyser, R. M. (1998). Beyond focus on form. Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & M. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42-63). Cambridge, UK: Cambridge University Press.

Doughty, C., & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197-261). Cambridge, UK: Cambridge University Press.

Edwards, C., & Willis, J. (Eds.) (2005). *Teachers exploring tasks in English language teaching*. Basingstoke, UK: Palgrave Macmillan.

Ellis, R. (2003). *Task-based language learning and teaching.* Oxford, UK: Oxford University Press.

Ellis, R. (2005). *Instructed second language acquisition: A literature review.* Wellington, NZ: Learning Media Ltd.

Foster, P., & Skehan, P. (1996). The influence of planning on performance in task-based learning. *Studies in Second Language Acquisition, 18*(3), 299-324.

Johnson, K. (1996). *Language teaching and skill learning.* Oxford, UK: Blackwell.

Kim, Y. (2013). Effects of pretask modelling on attention to form and question development. *TESOL Quarterly, 47*(1), 8-35.

Kim, Y., & McDonough, K. (2011). Using pretask modelling to encourage collaborative learning opportunities. *Language Teaching Research, 15*(2), 183-199.

Klapper, J. (2003). Taking communication to task? A critical review of recent trends in language teaching. *Language Learning Journal 27*, 33-42.

Kumaravadivelu, B. (1991). Language learning tasks: Teacher intention and learner interpretation. *ELT Journal, 45*(2), 98-107.

Larsen-Freeman, D. (2009). Teaching and testing grammar. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 518-542). Chichester, UK: Wiley-Blackwell.

Long, M. H. (1985). A role for instruction in second language acquisition: Task-based language teaching. In K. Hyltenstam and M. Pienemann (Eds.), *Modelling and assessing second language acquisition.* Clevedon, UK: Multilingual Matters.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). Amsterdam, NL: John Benjamin.

Long, M. H. (2015). *Second language acquisition and task-based language teaching.* Chichester, UK: Wiley Blackwell.

Long, B., & Kurzwell, J. (2002). PPP under the microscope. *English Teaching Professional 25*, 18-20.

Nunan, D. (2004). *Task-based language teaching.* Cambridge, UK: Cambridge University Press.

Nunan, D. (2007). Task-based language teaching in the Asia context: Defining 'task'. *Asian EFL Journal, 8*(3). Retrieved April 15, 2013 from http://www.asian-efl-journal.com/quarterly-journal/2006/09/29/task-based-language-teaching-in-the-asia-context-defining-task

Ortega, L. (2011). Sequences and processes in language learning. In M. H. Long & C. J. Doughty (Eds.), *Handbook of language teaching* (pp. 81-105). Malden, MA: Wiley-Blackwell.

Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communicative tasks for second language instruction. In G. Crookes, & S. M. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 9-34). Clevedon, UK: Multilingual Matters.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics 43*, 1-32.
Robinson, P. (Ed.) (2011). *Task-based language learning.* Malden, MA: Wiley-Blackwell.
Samuda, V., & Bygate, M. (2008). *Tasks in second language learning.* Basingstoke, UK: Palgrave Macmillan.
Sato, R. (2010). Reconsidering the effectiveness and suitability of P-P-P and TBLT in the Japanese EFL classroom. *JALT Journal 32*(2), 189-200.
Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics 11*(2), 129-158.
Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge, UK: Cambridge University Press.
Seedhouse, P. (1999). Task-based interaction. *ELT Journal, 53*(3), 149-156.
Seedhouse, P. (2004). *The interactional architecture of the language classroom: A conversation analysis perspective.* Malden, MA: Blackwell Publishing.
Sharwood Smith, M. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition, 15*, 165-179.
Skehan, P. (1996). Second language acquisition research and task-based instruction. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 17-30). Oxford, UK: Macmillan.
Skehan, P. (1998). *A cognitive approach to language learning*. Oxford, UK: Oxford University Press.
Skehan, P., & Foster, P. (2005). Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 193-216). Amsterdam, NL: John Benjamins.
Thornbury, S. (1997). Reformulation and reconstruction: Tasks that promote 'noticing'. *ELT Journal 51*(4), 326-335.
Willis, J. (1996). *A framework for task-based learning*. London, UK: Longman.
Willis, D., & Willis, J. (2007). *Doing task-based teaching.* Oxford, UK: Oxford University Press.

Appendix
Transcripts of the task models

Task 1: Cinema trip (decision making)

```
Rod: So Jim do you wanna see a movie this weekend?
Jim: Sure what do you wanna see?
Rod: Well I saw 007 Skyfall last weekend so I'd like to see something
     with a little less action this time.
Jim: Okay. How about Frankenweenie?
Rod: Ooh I'm not really into animation.
Jim: Ah okay hmm what about The Hobbit?
Rod: Ah The Hobbit! I heard that was quite good.
Jim: Yeah yeah it's um produced by Peter Jackson so it must be good.
     Do you wanna see it in 3D or 2D?
Rod: Hmm is it the same price?
Jim: Ah I think 3D is a little bit more expensive.
Rod: Ah oh 400 yen but that's okay. Ah yeah 3D sounds pretty cool.
Jim: Okay so er where?
Rod: Er the Umeda Toho cinema is pretty convenient for me. How about
     you?
Jim: Yeah, not too bad for me. Um alright what time?
Rod: Afternoon I think.
Jim: Afternoon?
Rod: After lunch.
Jim: After lunch? Well do you wanna- do you wanna grab some lunch
     first?
Rod: Ah that sounds great. Do you know any places around there?
Jim: Yeah there's a- there's a nice little um Italian restaurant
     just around the corner.
Rod: Ah I love Italian.
Jim: Brilliant okay so how about the er the two-twenty?
Rod: Two-twenty works for me.
Jim: Okay.
Rod: But where should we meet?
Jim: Tell you what let's meet outside the theatre at about one
     o'clock.
Rod: One o'clock and then from there we'll go to the restaurant.
Jim: Yeah.
Rod: Then after that ah sounds great.
Jim: Alright.
Rod: One o'clock outside the cinema what like in the lobby?
Jim: In the lobby yes.
Rod: Okay see you then.
Jim: Yeah looking forward to it bye.
Rod: Bye.
```

Task 2: Scheduling task (jigsaw)

Mark: Okay so we need to finish this project by next Tuesday.
Olive: Right.
Mark: So we need to organise a time to meet em so how is Thursday evening for you?
Olive: Actually I can't do Thursday evening I have to cook dinner.
Mark: Ah okay.
Olive: So yeah. I have some friends coming over for dinner I can't do that sorry.
Mark: Er okay.
Olive: How's Friday afternoon for you?
Mark: Friday afternoon ah actually this weekend we're going er we're going away for the weekend so I need to get ready. Um yeah we'll be back on... How is Sunday night?
Olive: Sunday night won't work at all. I have some family meeting that I gotta go to. Ah how about Monday morning?
Mark: Monday morning um Monday morning. Maybe I could do that before classes.
Olive: Okay.
Mark: Okay so Monday morning sounds good? Um.
Olive: What time's good for you on Monday morning? I have a yoga class from nine.
Mark: Okay what time do you finish yoga class?
Olive: I could meet you at ten thirty.
Mark: Okay let's do ten thirty.
Olive: Does that work? Okay.
Mark: At Starbucks?
Olive: Starbucks - the perfect place to meet.
Mark: Okay alright see you then.
Olive: Great see you Monday morning

Chinese EFL Learners' Writing Strategy Use in Reading-to-Write and Reading-Listening-Writing Tasks

Xinling Zhang*
Shanghai University, Shanghai, China

Yan Zhou
Guangdong University of Foreign Studies, Guangzhou, China

Siyu Zhang
Shanghai University, Shanghai, China

Abstract

This paper is an investigation into Chinese EFL learners writing strategy use in reading-to-write and reading-listening-writing tasks. The data of 120 high and low level Chinese EFL learners indicated that: 1) both tasks sampled the subjects' ability for using discourse synthesis strategy, self-monitoring strategy, and test-wiseness strategy, 2) L2 proficiency, task type, and their interaction partly influenced the type and amount of participants' use of the above strategies, 3) the high L2 proficiency group used selecting, organizing, and monitoring strategy more frequently and preferred evaluation strategy before writing in the reading-listening-writing task, 4) the low L2 proficiency group favored evaluation strategy in the post-writing phase in the reading-to-write task, 5) given the risk of plagiarism, the reading-listening-writing task was deemed more desirable.

Keywords: reading-to-write, reading-listening-writing, writing process

Introduction

"Integrated writing tasks are tasks in which test takers are presented with one or more language- rich source texts and are required to produce written compositions" (Knoch & Sitajalabhorn, 2013, p. 306). Many large-scale international high-stakes tests have been using integrated writing tasks, including TOEFL iBT, IELTS, TestDaF, DELE, the Russian Federal Foreign Russian Test, the new HSK test, and so on; this is also the case for some important English tests in China. In 2009, a large-scale high-stakes test, the National Matriculation English Test (NMET, college entrance examination) (Guangdong Province Version)[1] employed a reading-to-write item, requiring high school graduate candidates to read a passage before writing a summary and response on the same topic. In 2016, the Test for English Majors (TEM), a compulsory, standardized proficiency test for English majors in all Chinese universities, will launch its newest version in which integrated writing is used. This trend is also evident in the Public English Tests (PETS), of which the highest level is the language requirement for Chinese government scholarships for overseas study, and the list goes on.

* *Corresponding author. Tel: 00862166133062; Email: xlzhang_632@163.com; Mail Box No. 34, No.99, Shangda Road, Baoshan District, 200444, Shanghai, P.R. China*

Several authors advocate facilitating Chinese EFL learners reading via writing (e.g., Niu, 2014; Wang, Niu, & Zhen, 2000; Zhang & Zhou, 2014). However, research focusing on Chinese EFL learners' reading-listening-writing remains inadequate. Test scores are easily influenced by construct-irrelevant issues (Cumming, 2013; Esmaeili, 2002; Fox, 2003; Gebril & Plakans, 2013, 2014; Yu, 2013). The exploration of participants' writing processes or strategy use is considered to be of great importance to the construct validation of a test (Bachman, 2002; Knoch & Sitajalabhorn, 2013; Xu & Wu, 2012). The present study, taking reading-to-write and reading-listening-writing tasks as examples, explores the writing processes of Chinese EFL learners through integrated writing tasks, in the hope of providing evidence-based score explanations for all stakeholders.

Integrated Writing Tasks

Integrated writing tasks have various advantages: high authenticity (Hamp-Lyons & Kroll, 1996; Plakans, 2008), fairness (Weigle, 2002), avoiding background information bias, and providing positive washback (Xu & Gao, 2007). Integrated writing provides samples of the real-world academic writing process (Gebril & Plakans, 2013; Hamp-Lyons & Kroll, 1996; Yang, 2009), and represents the alignment between classroom writing teaching and EAP, thus activating language learners' motivation for learning (Plakans & Gebril, 2012). Moreover, Interdependence Hypothesis, Language Threshold Hypothesis, and the Extensive Reading Hypothesis (Grabe, 2003) hold that: 1) language learners' reading and writing abilities share a common core, 2) when their L2 proficiency reaches a threshold, their L1 literacy can be transferred to their L2 positively, 3) extensive reading could improve language learners' writing abilities (Flahive & Bailey, 1993).

Relevant assessment studies in the Chinese context have been conducted by Zhang et al. (2010), Zhang and Zeng (2009, 2010a, 2010b), Zhang and Zhou (2012, 2014), and He and Sun (2015). Zhang and Zhou (2014) found that Chinese EFL learners' writing scores differed across impromptu writing tasks, reading-to-write tasks, and reading-listening-writing tasks. The subjects achieved the highest mean score in the reading-listening-writing task; this task type has the additional benefit of eliminating potential plagiarism (Zhang & Zhou, 2014).

Studies on integrated English writing tasks in the Chinese context fall into the following categories: the facilitating effect of integrated writing on learners' reading, writing, and thinking (Xu & Gao, 2007) , the effectiveness of extensive reading on the enhancement of learners' writing abilities (Ji, 2009), and the characteristics of reading materials in varied reading-to-write tasks (Wu, 2014).

Specifically, the research available on reading-to-write tasks has focused on the following aspects: 1) writing products, textual features, and rating; 2) comparative studies of the writing products and processes between/among different types of reading-to-write tasks (summary, argumentative essay, etc.) (Asención, 2004; Zhang, 2009); and 3) writing process studies on reading-to-write tasks (Asención, 2004; Esmaeili, 2002; Plakans, 2008, 2009; Zhang, 2009). The major findings from the above studies include: 1) test takers are more likely to adopt connecting and organizing strategy than selecting; 2) test takers spend a significant amount of time monitoring, as they often resort to discourse synthesis strategy; and 3) how their strategy use in different integrated writing tasks (summary and reflective writing) is influenced by their L2 proficiency.

Yang (2009) and Yang and Plakans (2012) probed the strategy use of ESL learners enrolled in a large southwestern university in America. Yang and Plakans' (2012) participants were 161 non-native English speaking students from different countries including Brazil, China, Egypt, France, Iran, Japan, Korea, Mexico, Taiwan, Turkey, and Vietnam. They found that the participants' self-regulatory strategy use not only influenced their discourse synthesis strategy use and test-wiseness strategy use, but also their score and plagiarism strategy use. Also, a positive correlation was detected between patch writing strategy and test-wiseness strategy; whereas a negative correlation was detected between patch writing strategy and discourse synthesis strategy.

However, these studies have not addressed the following three aspects: first, none has displayed a comparative study on the strategy use between reading-to-write and reading-listening-writing tasks; second, the participants were mostly limited to EFL/ESL learners studying or working in English-speaking countries except Zhang and Zhou (2014); and third, none has taken the factor of L2 proficiency into consideration when comparing Chinese EFL learners' writing processes. Thus, this paper aims to investigate the effects of task type and L2 proficiency on Chinese EFL learners' strategy use in reading-to-write and reading-listening-writing tasks.

Strategies in the Writing Process

The present study delineates "process" as three nonlinear overlapping stages: the pre-writing, the during-writing, and the post-writing stages. As a compensation system that influences writing behavior intentionally, strategy is defined as writing skills used consciously so as to realize certain writing requirements, including discourse synthesis strategy, self-regulatory strategy, and test-wiseness strategy (Yang, 2009).

Discourse synthesis strategy represents a series of interactions among writers, source texts, and written texts (Yang, 2009, p. 47). It refers to the process of test takers' selection of relevant material from different resources, connecting them (Spivey & King, 1989), and then organizing them to produce a new structure. Self-regulatory strategy includes monitoring and evaluating (Yang, 2009, p. 48). Test-wiseness strategy involves test takers utilization of the formats or characteristics of a test or a test condition to compensate for their abilities (Yang, 2009, p. 49). In the present study, three types of test-wiseness strategy (using template, patch writing, and plagiarism) are employed. Using template refers to test takers' behavior of filling in ideas from the listening and/or reading material into a writing template they have memorized beforehand. Patch writing is the test takers' combination of their own ideas with the information from the listening and/or reading to form the content of their writing, while plagiarism is the test takers' verbatim use of the content of the listening and/or reading material when the task instructions clearly disallow it.

Pre-writing is the preparation work done before writing (Plakans, 2008; Yang, 2009), including selecting, connecting, and organizing in discourse synthesis strategy, plagiarism in test-wiseness strategy, and evaluation in self-regulatory strategy. Evaluation at this stage refers to test takers' predictions about the content of the source material. The during-writing stage involves first draft writing, which is comprised of monitoring in self-regulatory strategy, using writing template strategy, and patch strategy in test-wiseness strategy. Monitoring in the during-writing stage refers to examining the language, content, etc. of the written product. The post-writing stage ranges from finishing the first draft, through revision, to establishing a final version. Evaluation and monitoring in self-regulatory strategy are highlighted at the post-writing stage. Evaluation concerns the assessment of the written product, including the fulfillment of task requirements (Esmaeili, 2002), argument support, verbatim copy of the source materials, etc. Monitoring refers to the rereading of written products to examine whether there are grammatical mistakes, spelling mistakes, etc. (Asención, 2004; Plakans, 2009).

Writing strategy use is task-bound and context-specific (Glaser, 1984; Zhang, 2009). It is hypothesized that, other than test-wiseness strategy, the more frequently the above strategies are used, the better the result of a written product. Therefore, the present study addressed the following research questions:

1. Do Chinese EFL learners use discourse synthesis strategy, self-regulatory strategy, and test-wiseness strategy in reading-to-write and reading-listening-writing tasks?
2. Are there any differences in Chinese EFL learners' strategy use between reading-to-write and reading-listening-writing tasks? If yes, how?
3. Does L2 proficiency influence Chinese EFL learners' strategy use between reading-to-write and reading-listening-writing tasks? If yes, how?

Methodology

Study Design

Task type (reading-to-write task and reading-listening-writing task) and Chinese EFL learners' L2 proficiency (high level group and low level group) were the two independent variables in this study, while the subjects' writing strategy use was the dependent variable. Sixty master's degree candidates majoring in foreign linguistics and applied linguistics from a top international studies university in China (high level group) and 60 sophomore

English majors from an adult education institute (low level group)[2] were invited for the present study to finish both reading-to-write and reading-listening-writing tasks.

Table 1
Writing Strategy Framework

Task type	Writing stage	Strategy		Item	Content
Reading-listening-writing	pre-writing	discourse synthesis	connecting	1,2,4,5	Record important information and expressions from the reading and listening materials.
			organizing	6,7,8,10	Understand the similarities and differences between the reading and listening materials. Seek and understand the organization between sentences and paragraphs. Pay attention to the structure of the reading and listening materials to help understand them better.
			selecting	3,9	Select and summarize the ideas of the input materials based on comprehension.
		self-regulatory	evaluation	12	Make a prediction of the listening material after reading the reading passage.
		test-wiseness	plagiarism	11	Copy words and sentences directly from the reading and/or listening materials.
	during writing	self-regulatory	monitoring	13,16,17, 18,19	Check (reread) the written products to see whether there are direct copies from the input material, or grammatical mistakes, or expression mistakes, etc.
		test-wiseness	using template	14	Make use of the writing template in memory to finish the writing tasks.
			patch writing	15	Patch different sentences and content in the reading and/or listening materials.
	after writing	self-regulatory	evaluation	20, 24	Use examples to support the argument. Check so as to eliminate plagiarism.
			monitoring	21, 22, 23	Check (reread) to correct the linguistic mistakes (grammar, expression, and so on). Make sure that task fulfills the requirements (connecting listening and reading materials).
Reading-to-write	pre-writing	discourse synthesis	connecting	1,2	Selecting the main ideas from the reading passage. Make a writing plan.
			organizing	3,4,5,6	Seek and understand the organization between sentences and paragraphs. Pay attention to the structure of the reading material to help understand its content.
		test-wiseness	plagiarism	7	Copy words and sentences directly from the reading material.
	during writing	self-regulatory	monitoring	8,11,12, 13,14	Check (reread) the written products to see whether there are direct copies from the input material, or grammatical mistakes, or expression mistakes, etc.
		test-wiseness	using template	9	Make use of the writing template in memory to finish writing tasks.
			patch writing	10	Patch different sentences and content in the reading material.
	after writing	self-regulatory	evaluation	15,18	Use examples to support the argument. Check so as to eliminate plagiarism.
			monitoring	16,17	Check (reread) to correct the linguistic mistakes (grammar, expression, and so on). Make sure that the task fulfills the requirements (connecting listening and reading materials).

Instruments[3]

The instruments adopted in the present study were 2 writing tasks (reading-to-write task and reading-listening-write task) and 2 questionnaires. Both tasks took Grade Inflation as the topic. The reading-to-write task required the students to read a 300 word English passage (about 5 minutes) before they finish a 200-300 word argumentative essay in 30 minutes. The other task asked the students to first read a 300 word English passage (about 5 minutes), then listen to an English video clip of about 300 words (about 2-3 minutes), and finish a 200-300 word argumentative essay in 30 minutes. The vocabulary of the reading and listening materials was either from the College English Test Band 4 (CET-4) vocabulary list or a Chinese equivalent. The reading speed required was about 55 words per minute and the broadcasting speed of the listening material was around 125 words per minute. Both are lower than the requirements of CET-4. The reading passage, writing task requirements, and a brief rating criterion were attached to the test paper for both writing tasks.

Two questionnaires, adapted from Yang (2009), were used in the present study to investigate test takers' strategy use in reading-to-write and reading-listening-writing tasks, respectively. Each questionnaire consisted of a set of Likert scale items and one semi-open question. Table 1 reports the constructs of each questionnaire, the corresponding items of each strategy, as well as the content each strategy examined. The semi-open question was aimed at detecting the difficulties that the test takers encountered in reading, listening, and writing by the test takers.

Data Collection

The simulation test followed the procedures of a formal examination. Each group of participants finished the corresponding questionnaire items right after finishing the writing task. To ensure their accuracy, the data collected from questionnaires were typed into Excel sheets by the second author, and proofread by the other two authors. Statistical analyses were conducted via SPSS 21.0.

Analytical Procedure

Reliability of the reading-listening-writing and reading-to-write questionnaires was tested with Cronbach's α values of 0.79 and 0.83, respectively. Exploratory factor analysis was performed on the data from the five-point Likert scale items as a means of validating the questionnaires. The KMO and significance of Barlett's Test of Sphericity for the reading-to-write and the reading-listening-writing were .755 and .793 (> .05), is .00 and .01 (<.05), respectively. The three factors extracted (eigenvalue >1) were named discourse synthesis strategy, self-regulatory strategy, and test-wiseness strategy for both questionnaires in light of the aforementioned construct. We also invited two linguistics experts to read the open-ended questionnaire items. The responses from the experts justified the validity requirement. The data satisfied the prerequisites for ANOVAs (Tables 3, 5, and 7) in terms of the independence of the samples, the normality of the data distribution, and p values (>.05) for Levene's test.

Results

Pre-writing Stage

Overall, about 80-98% (Table 2) of the students made use of the three types of discourse synthesis strategy and evaluation in the pre-writing stage in both tasks.

Task type and L2 proficiency partially exerted influence on the participants' use of discourse synthesis strategy, self-regulatory strategy, and test-wiseness strategy. In the case of connecting, descriptive statistics showed that there was no great variation in the values of mean and standard deviation between the two proficiency groups across the two tasks (Table 2).

Results of the ANOVA indicated that task type and L2 proficiency did not have significant influence on high and low level subjects' connecting strategy use in the pre-writing stage (Table 3).

Table 2
Descriptive Statistics of High and Low Level Participants' Strategy Use in the Pre-Writing Stage

L2	Task type	Strategy		1(%)	2(%)	3(%)	4(%)	5(%)	Use (%)	Mean/SD
high	reading-listening-writing	discourse synthesis	connecting	10.00	15.00	30.83	29.17	15.00	90.00	3.24/1.18
			organizing	10.83	20.00	30.83	32.50	5.83	89.17	3.02/1.09
			selecting	1.67	10.00	13.33	50.00	25.00	98.33	3.87/.97
		test-wiseness	plagiarism	16.67	40.00	16.67	13.33	13.33	83.33	2.67/1.30
		self-regulatory	evaluation	10.00	23.33	43.33	16.67	6.67	90.00	2.87/1.04
	reading-to-write	discourse synthesis strategy	connecting	18.33	21.67	15.00	25.00	20.00	81.67	3.07/1.42
			organizing	8.33	10.83	28.33	34.17	18.33	91.67	3.43/1.16
		test-wiseness	plagiarism	13.33	20.00	30.00	20.00	16.67	86.67	3.07/1.29
low	reading-listening-writing	discourse synthesis	connecting	3.33	18.33	29.17	32.50	16.67	96.67	3.41/1.07
			organizing	2.50	20.00	39.17	33.33	5.00	97.50	3.18/.92
			selecting	3.33	15.00	33.33	28.33	20.00	96.67	3.47/1.08
		test-wiseness	plagiarism	20.00	23.33	23.33	30.00	3.33	80.00	2.70/1.18
		self-regulatory	evaluation	3.33	13.33	26.67	43.33	13.33	96.67	3.50/1.01
	reading-to-write	discourse synthesis	connecting	8.33	23.33	25.00	35.00	8.33	91.67	3.12/1.12
			organizing	4.17	26.67	25.00	40.83	3.33	95.83	3.13/.98
		test-wiseness	plagiarism	6.67	13.33	43.33	16.67	20.00	93.33	3.30/1.15

Note. 1= never, 2=seldom, 3=occasionally, 4=often, 5=always

As for organizing, descriptive statistics were higher in the reading-listening-writing task (3.02, 1.09) than those in the reading-to-write task (3.43, 1.16, 80.8%) for the high L2 group (Table 2). The high (3.43, 1.16) and low-level (3.18, .92) groups' means and standard deviation values differed in the reading-to-write task. ANOVA tests indicated that the two independent variables did not exert significant influence on the participants' organizing strategy use (Table 3); however, their interaction did ($F=6.29$, $p<.05$). Simple main effect analysis[4] showed that the high level participants used organizing strategy more frequently in the reading-to-write task than in the reading-listening-writing task ($F=9.63$, $p<.05$). They also used it more often than the low level participants in the reading-to-write task ($F=5.27$, $p<.05$).

In selecting, descriptive data of the high level participants' in the reading-listening-writing task were higher than those of the low level participants (Table 2). ANOVA results demonstrated that L2 proficiency significantly influenced participants' selecting strategy use in the pre-writing stage ($F=4.57$, $p<.05$).

Evaluation at this stage was used less frequently by the high level participants than their lower level counterparts (Table 2) in reading-listening-writing. L2 proficiency had an influence on participants' evaluation strategy use ($F=5.72$, $p<.05$) (Table 3).

Both proficiency groups used plagiarism strategy less frequently in the reading-listening writing task than in the reading-to-write task (Table 2). Task type exerted statistically significant influence on the participants' plagiarism strategy use in the pre-writing stage ($F=4.97$, $p<.05$) (Table 3).

Table 3
ANOVA Results of High and Low Level Subjects' Strategy Use in the Pre-Writing Stage

Strategy		Source of difference	F value	Significance level
Discourse synthesis	connecting	proficiency	.67	.41
		task type	3.12	.08
		proficiency* task type	.20	.66
	organizing	proficiency	.56	.46
		task type	3.58	.06
		proficiency*task type	6.29	.01
	selecting	proficiency	4.57	.04
Self-regulatory	evaluation	proficiency	5.72	.02
Test-wiseness	plagiarism	proficiency	.35	.55
		task type	4.97	.03
		proficiency*task type	.20	.66

During-writing Stage

The percentage of participants using monitoring, template, and patch writing strategy in both task types ranged from 86.67% to 97.33% (Table 4).

Task type did not have significant influence on the participants' template using, plagiarism, or monitoring strategy (Table 5), but L2 proficiency did. To be specific: 1) the high level participants' values of mean and standard deviation in reading-to-write (3.63, 1.23) and reading-listening-writing (3.47, 1.25) tasks were higher than those of the low level participants (reading-to-write: 3.16, 1.10; reading-listening-writing: 3.27, .98). ANOVA detected that L2 proficiency significantly influenced the participants' monitoring strategy use during the writing stage ($F=12.76$, $p<.05$) (Table 5). 2) L2 proficiency exerted significant influence on patch writing strategy use ($F=4.77$, $p<.05$) (Tables 4, 5). 3) Little difference was found between the two groups' template using strategy in both tasks (Tables 4 and 5).

Table 4
Descriptive Statistics of High and Low Level Participants in the During Writing Stage

L2	task type	strategy type		1(%)	2(%)	3(%)	4(%)	5(%)	Use (%)	Mean/SD
High	reading-listening-writing	self-regulatory	monitoring	10.00	13.33	17.33	38.00	21.33	90.00	3.47/1.25
		test-wiseness	using template	6.67	26.67	16.67	20.00	30.00	93.33	3.40/1.35
			patch writing	10.00	46.67	20.00	10.00	13.33	90.00	2.70/1.21
	reading-to-write	self-regulatory	monitoring	8.00	10.67	20.67	32.00	28.67	92.00	3.63/1.23
		test-wiseness	using template	13.33	6.67	26.67	33.33	20.00	86.67	3.37/1.27
			patch writing	13.33	23.33	30.00	20.00	13.33	86.67	2.97/1.25
Low	reading-listening-writing	self-regulatory	monitoring	2.67	18.00	40.00	28.00	11.33	97.33	3.27/.98
		test-wiseness	using template	0.00	16.67	30.00	36.67	16.67	100.0	3.53/.97
			patch writing	10.00	13.33	40.00	23.33	13.33	90.00	3.17/1.15
	reading-to-write	self-regulatory	monitoring	9.33	14.67	37.33	28.00	10.67	90.67	3.16/1.10
		test-wiseness	using template	6.67	3.33	30.00	50.00	10.00	93.33	3.53/.97
			patch writing	3.33	20.00	20.00	43.33	13.33	96.67	3.43/1.07

Note. 1= never, 2=seldom, 3=occasionally, 4=often, 5=always

Table 5
ANOVA Results of High and Low Level Subjects' Strategy Use in the During Writing Stage

Strategy		source of difference	F value	significance level
self-regulatory	monitoring	proficiency	12.76	.00
		task type	.05	.83
		proficiency*task type	2.04	.15
test-wiseness	using template	proficiency	.51	.48
		task type	.01	.94
		proficiency*task type	.01	.94
	patch writing	proficiency	4.77	.03
		task type	1.56	.21
		proficiency*task type	.00	1.00

Post-writing Stage

About 87-99% of all the participants used evaluation and monitoring strategy in the post-writing stage in both tasks (Table 6). Results of ANOVA indicated that task type and L2 proficiency did not have statistical influence on evaluation strategy use (task type: F=.01, p>.05; L2 proficiency: F=1.16, p>.05) (Table 7), but the marginal interactive influence (F=3.67, p=.06) (Table 7) might indicate that higher level participants tended to use evaluation in the reading-to-write task.

The high level participants' descriptive data values in monitoring strategy use were higher than those of the low level participants in both tasks (Table 6). ANOVA and simple main effect analysis detected that task type and L2 proficiency had an interactive influence on monitoring in the post-writing stage: 1) the high level participants' used monitoring more frequently than the other group in the reading-to-write task (F=4.58, p<.05); 2), they also used monitoring more frequently in the reading-to-write task compared to the reading-listening-writing task (F=3.58, p<.05) (Table 7).

Table 6
Descriptive Statistics of High and Low Level Participants in the Post-Writing Stage

L2	task type	strategy		1 (%)	2 (%)	3(%)	4(%)	5(%)	Use (%)	Mean/SD
high	reading-listening-writing	self-regulatory	evaluation	11.67	20.00	28.33	25.00	15.00	88.33	3.12/1.24
			monitoring	12.22	20.00	20.00	25.56	22.22	87.78	3.26/1.34
	reading-to-write	self-regulatory	evaluation	8.33	8.33	36.67	28.33	18.33	91.67	3.40/1.14
			monitoring	3.33	11.67	16.67	35.00	33.33	96.67	3.83/1.12
low	reading-listening-writing	self-regulatory	evaluation	1.67	20.00	41.67	26.67	10.00	98.33	3.23/.95
			monitoring	4.44	23.33	28.89	27.78	15.56	95.56	3.27/1.12
	reading-to-write	self-regulatory	evaluation	5.00	26.67	38.33	25.00	5.00	95.00	2.98/.97
			monitoring	6.67	18.33	40.00	23.33	11.67	93.33	3.15/1.07

Note. 1= never, 2=seldom, 3=occasionally, 4=often, 5=always

Table 7
ANOVA Results of High and Low Level Subjects' Strategy Use in the Post-Writing Stage

Strategy		source of difference	F value	significance level
self-regulatory	evaluation	proficiency	1.16	.28
		task type	.01	.91
		proficiency*task type	3.67	.06
	monitoring	proficiency	5.83	.02
		task type	2.75	.10
		proficiency*task type	6.23	.01

Discussion

Discussions of the above findings will be conducted from the perspectives of writing process model (Hayes, 1990), cognitive theories of multimedia learning (Mayer & Moreno, 2003), and previous studies on integrated writing tasks. First, the results of this study are summarized as follows (Table 8):

Table 8
A Summary of the Influences of Independent Variables

Strategy		Task type	L2	Interaction
Discourse synthesis strategy	selecting	×	√	×
	connecting	/	/	/
	organizing	/	/	√
Self-regulatory strategy	/×	√√	/√	
Evaluation (pre-writing writing stages in reading-listening-writing task/ post-writing stage in reading-to-write and reading-listening-writing tasks)	××	√√	×√	
Test-wiseness strategy	plagiarism	√	×	×
	using template	×	×	×
	patch writing	×	√	×

Constructs of Reading-to-write and Reading-listening-writing

The majority of participants used discourse synthesis strategy, self-regulatory strategy, and test-wiseness strategy in both tasks in the three writing stages, indicating that both tasks might have sampled test takers' ability of retrieving, processing, and organizing multi-modal information in the target language to complete the assigned writing tasks, which defines the construct of the test, echoing Plakans' (2009) findings. In this process, the participants made use of verbal information processing and visual information processing systems to process and transform source materials and task requirements (Mayer, 1996). They selected key information from the source materials, constructed situational models, organized new information, and integrated it into their previous knowledge using knowledge of the topic, target reader, genre, etc. in their working memory (Sweller, 1988). They coordinated the internal cognitive loads in integrated writing tasks and their components, and the external cognitive loads accompanied by the source materials. To be exact, they probed into the relationships between sentences and paragraphs, made writing plans, supported their arguments by using multi-modal information, and revised mistakes.

Task Type Influence

Task type had little influence on discourse synthesis strategy and self-regulatory strategy use, which might be explained as below. According to Hayes' writing process model (1996), task environment and language learners' individual factors influence the writing process. Selecting, connecting, and organizing in discourse synthesis strategy are inherently associated with the task characteristics in Hayes' (1996) writing process model. When task characteristics are retrieved from test takers' working memory, the strategies activated would be relatively stable and predictable. Similarly, self-regulatory strategy is fundamentally connected with meta-cognitive strategies (Spivey, 1990), L2 proficiency and individual factors; the possibility that they are influenced by task type is relatively low.

Similar to the results of Zhang and Zhou (2014), task type exerted significant influence on participants' plagiarism in test-wiseness strategy in the pre-writing stage only (see Section 4.1), which may indicate that bi-modality input can effectively reduce plagiarism compared to mono-modality input. As plagiarism has an obvious negative influence on language learners' writing performance (Yang, 2009; Yang, & Plakans, 2012), a reading-listening-writing task could tap students' writing ability better when the listening input has an appropriate level of difficulty.

L2 Proficiency Influence

L2 proficiency exerted restricted influence on some participants' discourse synthesis strategy, self-regulatory strategy, or test-wiseness strategy use. The high L2 level group used selecting strategy more frequently. They recorded and summarized the main idea of the reading or listening material in the pre-writing stage better, and did the checking and avoiding of plagiarism in the post-writing stage more successfully. This could be an indication that they were in a better position to fulfill the mission concerning task instructions, source materials, and task requirements in opposition to the low L2 level group. We might tentatively conclude that the higher the L2 level of the test takers, the better they were able to make use of the strategies, and the writing they produced was also superior in language, content, and structure (Zhang & Zhou, 2014). However, this could also be an indicator that integrated writing testing tasks are capable of differentiating between candidates of different L2 levels.

Meanwhile, the interactive influence of task type and L2 proficiency on the participants' organizing strategy use in the pre-writing stage might be attributed to the high level participants' abundant use of organizing in reading-to-write. They made more efforts to integrate and organize the selected information, which verified the findings of Tomas (2011) and Wette (2010). In the post-writing stage, the interactive influence of task type and L2 proficiency was significant on monitoring strategy use (see Section 4.3), which could be attributed to the high L2 level group as well. This is possibly due to the reason that they borrowed little information from the reading/listening material, and produced more on their own, which in turn required more attention to linguistic mistake correction and expression modification.

Notably, most of the low level participants (96.7%) had difficulty in listening. This might help explain why many of them used evaluation strategy in the pre-writing stage because prediction of the listening material could help improve their listening outcomes in the pre-writing stage; besides, they were less aware of using monitoring strategy. Participants tended to use patch writing strategy by blending the content of source materials instead of integrating and organizing the source materials, which possibly could be attributed to task familiarity, time pressure from bi-modality input in the reading-listening-writing task, and limited information processing capacity. For the low level participants, the dual-channel processing capacity (Mayer & Moreno, 2003) needs improvement, the ability to capture listening information in particular.

Lastly, the relatively low frequency of participants' use of organizing strategy could indicate that the bi-modality input in the reading-listening-writing task was difficult for all participants. This can most probably be attributed to the priority given to selecting strategy and connecting strategy use over organizing strategy use when participants' information processing ability and attention resources are limited, due to the high-level nature of organizing (Weigle, 2002) to yield new structures (Spivey & King, 1989).

Conclusion

The findings of this study can be summarized as follows:

First, the reading-to-write and the reading-listening-writing tasks have a similar strategy construct framework and internal stability. Both tasks tap participants' discourse synthesis strategy, self-regulatory strategy, and test-wiseness strategy use, but may differ in dimensions.

Second, L2 proficiency has a greater effect on the participants' writing performance, compared to task type. Studies show that positive relations exist between L2 proficiency level and second language writing ability (Grabe, 2003; Ma, 2004), demonstrating that on the premise of avoiding plagiarism and appropriate source material choice.

Third, the bi-modality input on the reading-listening-writing task can effectively avoid plagiarism of the source material compared to the mono-modal input in the reading-to-write task.

The following limitations might affect the overall generalizability of the conclusions of the present research. More qualitative data gained via thinking aloud and interviews may reveal the thinking processes of the participants in the two conditions; in addition, the sample size could also be enlarged. However, the present research is implicational in the Chinese English L2 context which shares the biggest English testing population worldwide in terms of the following considerations. To begin with, we would argue that integrated writing is an effective language testing device only when a candidates' L2 reaches a certain threshold as it requires multi-dimensional ability and language testing is a means instead of the purpose for language instruction. In addition, test item writers should be cautious about the difficulty of the reading and listening passages used in integrated writing tasks. Furthermore, taking into consideration that multi-modal communicative language competence is demonstrated in more performance tests worldwide, we would argue, together with Paz and Felton (2010) that more emphasis be put on integrated English writing in the Chinese context, in the firm belief that Chinese EFL learners' faster, effective, and accurate extraction of information from various sources could facilitate them to produce richer, more elaborate argumentative essays.

Endnotes

[1] The National Matriculation English Test (NMET) (Guangdong Province Version) is developed on a provincial basis in Guangdong Province, though most provinces use one single national test paper for English in the college entrance examination in China.

[2] Juffs' (1996) criterion of formal learning time was adopted to distinguish between high and low level proficiency groups in the present study. Students enrolled in Chinese Adult Education classes are mostly much poorer than normal English majors as they either failed in National Matriculation English Test for college entrance, or major in other fields. They only study on weekends and one night a week. We would argue that the miscellaneous Chinese EFL learners are representative enough for an experimental study.

[3] Omitted, available on request.

[4] Omitted, available on request.

References

Asención, Y. (2004). *Validation of reading-to-write assessment tasks performed by second language learners* (Unpublished Doctoral Dissertation). Northern Arizona University, Flagstaff.

Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19 (4), 453–476.

Chamot, A. U. (1987). The learning strategies of ESL students. In A. L. Wenden & J. Rubin (Eds.). *Learner strategies in language learning* (pp. 71-84). NJ, America: Prentice Hall.

Cumming, A. (2013). Assessing integrated tasks for academic purposes: promises and perils. *Language Assessment Quarterly* 10(1), 1-18.

Esmaeili, H. (2002). Integrated reading and writing tasks and ESL students' reading and writing performance in an English language test. *The Canadian Modern Language Review*, 58 (4), 599-622.

Flahive, D., & Bailey, N. (1993). Exploring reading/writing relationships in adult second language learners. In Carson, J. & Leki, I. (Eds.) *Reading in the composition classroom: Second language perspectives* (pp. 128-140). Boston: Heinle & Heinle Pub.

Fox, J. (2003). From products to process: An ecological approach to bias detection. *International Journal of Testing*, 3(1), 21–47.

Gebril, A., & Plakans, L. (2013). Towards a transparent construct of reading-to-write tasks: The interface between discourse features and proficiency. *Language Assessment Quarterly*, 10(1), 9-27.

Gebril, A., & Plakans, L. (2014). Assembling validity evidence for assessing academic writing: Rater reactions to integrated tasks. *Assessing Writing*, 21(1), 56-73.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist*, 39 (1), 93-104.

Grabe, W. (2003). *Reading and writing relations: second language perspectives on writing* (pp. 242-262). Cambridge, UK: Cambridge University Press.
Hamp-Lyons, L., & Kroll, B. (1996). Issues in ESL writing assessment: An overview. *College ESL*, 6 (1), 52-72.
Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & R. S (Eds.), *The science of writing* (pp. 29-55). NJ, America: Lawrence Erlbaum.
He, L., & Sun. Y. (2015). Investigation of the effects of prompt characteristics on Chinese test-takers' integrated writing performance. *Foreign Language Teaching and Research*, 47 (2), 237-250.
Ji, X. (2009). The effects of extensive writing and reading-writing integration on writing proficiency. *Journal of Tianjin Foreign Studies University*, 16 (5), 65-71.
Knoch, U., & Sitajalabhorn, W. (2013). A closer look at integrated writing tasks: Towards a more focused definition for assessment purposes. *Assessing Writing,* 18(4), 300-308.
Mayer, R. E. (1996). Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review*, 8 (4), 357-371.
Mayer. R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 43-52.
Paz, S. D. L., & Felton, K. M. (2010). Reading and writing from multiple source document in history: Effects of strategy instruction with low to average high school writers. *Contemporary Education Psychology,* 35 (1), 174-192.
Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, 13 (1), 111–129.
Plakans, L. (2009). Discourse synthesis in integrated second language writing assessment. *Language Testing*, 26 (4), 561-587.
Plakans, L., & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. *Assessing Writing*, 17 (1), 18 – 34.
Spivey, N. (1990). Transforming texts constructive processes in reading and writing. *Written Communication*, 7(2), 256-287.
Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly*, 24(1), 7-26.
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning, *cognitive science*, 12(2), 257–285
Tomas, Z. (2011). *Textual borrowing across academic assignments: Examining undergraduate second language writers' implementation of writing instruction* (Unpublished Doctoral Dissertation). The University of Utah, Salt Lake City.
Weigle, S. (2002). *Assessing writing*. New York, America: Cambridge University Press.
Wette, R. (2010). Evaluating student learning in a university-level EAP unit on writing using sources. *Journal of Second Language Writing*, 19 (3), 158-177.
Xu, Y., & Wu, Z. (2012). Test-taking strategies for a high-stakes writing test: An exploratory study of L2 Chinese EFL learners. *Assessing Writing*, 17 (3), 174-190.
Yang, H. (2009). *Exploring the complexity of second language writers' strategy use and performance on an integrated writing test through structural equation modeling and qualitative approaches* (Unpublished Doctoral Dissertation). The University of Texas at Austin, Austin.
Yang, H., & Plakans, L. (2012). Second language writers' strategy use and performance on an integrated reading-listening-writing task. *TESOL Quarterly*, 46 (1), 80-103.
Yu, G. (2013). From integrative to integrated language assessment: Are we there yet? *Language Assessment Quarterly*, 10(1), 110-114.
Ma, X. (2004). *The quantitative analysis of reading-writing relationships in second language* (Unpublished Master Thesis). Beijing Language and Culture University, Beijing.
Niu, J. (2014). *Exploration of integrated writing in language tests ——On "Listening and Writing" in IBCET4* (Unpublished Master Thesis). East China University of Science and Technology, Shanghai.

Wang, Ch., Niu, R., & Zhen, X. (2000). Improving English through writing. *Foreign Language Teaching and Research*. 32(3), 207-212.

Wu, Y. (2014). Exploring test takers' source use strategies in an EFL reading-to-write integrated task. *Computer Assisted Foreign Language Education*, 159 (9), 63-69.

Xu, H, & Gao, C. (2007). An empirical study on integrating reading with writing in EFL teaching. *Modern Foreign Languages,* 30(2), 184-190.

Zhang, X. (2009). *A construct validation research on text-based English writing task in China: A case study of NMET (GD)*. Shanghai, China: Shanghai University Press.

Zhang, X., Li, Q., & Zhang, J. (2010). A study of linguistic characteristics and score predictors for different reading-to-write tasks, *Modern Foreign Languages*, 33 (3), 287-296.

Zhang, X., & Zeng, Y. (2010). Task-based writing in NMET Guangdong version——an evidence-based orientation. *Journal of Guangdong University of Foreign Studies*, 21(2), 26-29.

Zhang, X., & Zeng, Y. (2010). Validation of a large-scale reading-to-write task： Evidence from multi-faceted research model analysis, *Journal of PLA University of Foreign Languages*, 33 (2), 50-54.

Zhang, X., & Zeng, Y. (2009). Construct validation of reading-to-write tasks in large-scale writing assessment in China, *Journal of PLA University of Foreign Languages*, 32 (1), 56-61.

Zhang, X. (2009). An overview of research on reading-to-write testing tasks, *Journal of Tianjin Foreign Studies University*, 16 (1), 75-80.

Zhang, X., & Zhou, Y. (2012). A review of research on integrative listening/reading-writing tasks. *Foreign Language Testing and Teaching*, (1), 21-26.

Zhang, X., & Zhou, Y. (2014). The effect of task type on Chinese EFL learner's writing performance. *Modern Foreign Languages,* 37(4), 548-558.

Coherence and Organization in Narrative Discourse by Intermediate EFL Writers

Salah Alfarwan*
Majma'ah University, Kingdom of Saudi Arabia

Abstract

Past research is divided as to whether L1 interference has a strong impact on the rhetorical (organizational) and linguistic (grammatical and lexical) features of EFL writing across proficiency levels. In order to illuminate this, 14 narrative compositions of intermediate male English major students from Saudi Arabia were scored and analyzed at multiple levels. Special attention was paid to the analysis of cohesive devices, coherence relations, narrative structure, and topic progression. Findings supported some previous work and revealed a wide range of occurrence of Arabic (L1) interference at the levels of lexis, syntax, and organization. Weak writers produced texts with seriously problematic cohesion and coherence and often the compositions of strong writers exhibited the same kinds of error, though in a less extreme form. Implications are discussed and recommendations made to improve writing skills among the students of English, by means such as: exposing EFL students to more authentic EFL texts written by L1 natives, introducing discourse analysis theories and practices to EFL instructors, and engaging students in more L2 contact with western materials and media.

Keywords: coherence, cohesion, composition, discourse analysis, EFL, interference, narrative

Introduction

Writing is a key language skill yet many Arabic learners exhibit essay composition weaknesses. One well-established applied linguistic approach to understanding the difficulties learners face is based on analyzing written texts (Hedge, 1988). From such studies, Arab students have often been found to transfer native language forms, phrasing, text organization, and even thoughts into their EFL writing (Connor, 1996; Kaplan, 1966, 1972; Rammel, 2009).

Kaplan's seminal 1966 study was particularly concerned with the transfer of first language cultural conventions into second language performance and dealt specifically with the rhetorical organization of ideas in writing, which was widely assumed to be culturally determined. Since Arabic written discourse is characterized by parallel structures and a widespread use of coordination, Kaplan captured in a famous diagram of Semitic paragraph organization how he believed EFL writers equally exhibit parallelism and clause coordination in their compositions. It will be seen through the current study whether what Kaplan argued is in fact true of EFL students' composition writing in our context.

This Study

This research project was prompted by Kaplan's (1966) controversial contrastive rhetoric hypothesis which suggested that EFL students transfer rhetorical patterns from their L1 to their EFL composition writing and that this is a characteristic of both weak and proficient EFL writers' compositions. As we will see below, the literature

* *Tel: 00966542884242; E-mail: S.alfarwan@mu.edu.sa; Address: English Department, Majma'ah University, College of Science and Humanities at Rumah, Majma'ah University, Kingdom of Saudi Arabia*

tends to support this hypothesis but the evidence is not unanimous. This topic is important because "use of rhetorical patterns unfamiliar to the intended audience … may not only strike readers as lack of rhetorical elegance, but as lack of coherent writing or even thinking, which can seriously affect the credibility of non-native writers" (Mauranen, 1993, p. 2). Lack of cohesion constitutes a serious problem in the writing of many EFL learners, as has been recognised by teachers and researchers alike (Halliday, 1994). It may lead to EFL writers not meeting the rhetorical expectations of a native speaker audience and to low marks from their teachers even if they write passable content (Currie, 1990).

Indeed, the English compositions written by Saudi English majors in the researcher's context do not achieve a satisfactory level of cohesion. In addition, the researcher is not aware of any relevant studies examining the coherence and narrative discourse of intermediate EFL writers in the current setting. For the above theoretical and pedagogical reasons, the researcher decided to carry out this small-scale study of Saudi students' coherence and high level organization when writing narrative discourse, addressing the following research questions:

1. What are the prominent non-native-like features of the written English narrative compositions of Saudi English major first year university students (of intermediate proficiency)? Especially, what do we find at the higher levels of rhetorical structure/text organization?
2. How much, and in what specific areas do strong and weak writers differ in these respects?
3. Is the reason for non-native-like performance by these writers primarily transfer from L1 Arabic or other factors?

Contrastive Rhetoric

Research into L2 writing began in the 1960s and early 1970s in the USA and the UK. It stemmed initially from L1 composition research, employing similar theories and types of analysis (Johnson & Roen, 1989; Kroll, 1990; Raimes, 1991; Reid, 1993). One early focus of attention in L2 writing research was the rhetorical variation noticed in the writing of L2 students from various L1 origins (Kaplan, 1966, 1972). Even EFL learners who are proficient in their first language may encounter "confusing differences between the conventions of writing in their first language and English" (Hedge, 1988, p. 5). This line of research, commonly referred to as contrastive rhetoric, drew on discourse analysis and linguistic text research to explore how students' writing could be analyzed at the discourse level as a way of understanding the organizational preferences in student writing. Researchers thus became aware of the importance of higher level features in EFL writing, including coherence and cohesion, and other devices this study aims to examine.

Within this field, the study of Arabic-English differences has been a specific area of interest. In Kaplan's (1966) early discussion of contrastive rhetoric, Arab students writing in English were seen as writing in a pattern characterized by repetition and elaborate parallelism, typical of Arabic, rather than in a linear pattern, as in English. Ostler (1987), in an effort to support Kaplan's (1966) generalization based on a small corpus of compositions, found that quite advanced Arab EFL students exhibited greater use of coordination and opposition than subordination, as well as greater use of parallelism in and between clauses. Ostler (1987) argued that Arabic EFL writers are heavily influenced by classical Arabic, a language which is more reflective of oral language traditions. Saadeddin (1989) further explained how Arab writers typically prefer an 'oral' mode of writing because it indicates solidarity and shared cultural beliefs. Its salient features include a free pattern of organization, frequent use of emphasis, repetition of specific syntactic structures, and development by addition and accumulation.

Smail (2010) more recently compared the EFL and Arabic L1 writing of 30 native Arabic speakers and the English L1 writing of 30 native English speakers on the same persuasive writing task. The researcher also administered a language history questionnaire to ascertain the participants' demographics and educational background. The findings threw serious doubt on the validity of the contrastive rhetoric hypothesis and suggested that other variables, individual, contextual, and/or situational, play a more significant role in a writers' rhetorical performance. By contrast, Stapa and Irtaimeh's (2012) research on the transfer of Arabic rhetorical features into English by 50 Jordanian EFL students once again supported the hypothesis that rhetorical features are transferred intensively from Arabic into English. Hence, it remains unclear how far high level transfer exists among Arab L1 writers, and how far it depends on proficiency level. We turn now to the various frameworks that have been proposed within which to analyse rhetorical structure and higher level organizational features of written discourse. Due to space considerations we limit the review to those that are used in our own study.

Cohesion and Coherence

Cohesion, as defined by Halliday and Hasan (1976), is the overt linguistic basis whereby a text is logically and sequentially organized. They later claimed that cohesion and coherence together constitute the two most important textual features of good writing (Halliday & Hasan, 1989). In Halliday and Hasan's (1976) well known framework, which we have adopted, cohesion is classified into five categories: Reference (pronominal, demonstrative, definite article *the*, and comparative); Substitution, *one/some/ones* (as substitutes of noun phrases), *do/so/it/that* (as substitutes of predicates), and *so/not* (as substitutes of clauses); Ellipsis (of noun phrases, the predicate, and a clause); Conjunction (additive, adversative, causative, temporal, and continuative); and Lexical (involving semantically related vocabulary).

A large number of studies have been carried out using Halliday and Hasan's (1976) framework, and a variety of contradictory results have been reported. Zhang (2000), for instance, found that there was no difference in the appropriacy of use of cohesion and coherence devices in higher and lower level compositions. However, Witte and Faigley (1981) found that good essays included more cohesive devices than poor essays, leading to a better quality end-product. In Bahrain, Alissa (2013) carried out a study with a group of 18 Arabic EFL learners and concluded that strong learners showed a better employment of cohesive devices and so produced better written compositions than weak learners who demonstrated both more difficulties using such advices and a clear impact of their L1.

While cohesion is defined as linguistic, coherence is not a matter of language at all, but of thought, though the connection of thought may of course be marked overtly by cohesive devices. A text is coherent because the later parts of the text make sense in the context of the earlier parts, not just because of the use or misuse of any particular linguistic 'device'. As Connor (1996) noted, "to be coherent, text needs to make sense to the reader" (p. 84). In other words, the meanings of different parts of a text (often referred to as propositions) have to be recognizably interrelated. Mann and Thompson (1986) pointed out that propositional relations are self-determining of vocabulary use to signal such relations. Propositional relations function as implicit mechanisms by which coherence is achieved.

There are many approaches to analysing coherence in terms of how propositions come to form a coherent text. Mann and Thompson (1986) proposed that there is a limited set of 15 rhetorical relations which can be used to characterize the semantic relations between parts of a text or propositions. These relations are not intended to be seen as labels that identify the functions of sentences in the text; rather, each relation can be thought to be supplying propositions with the quality of making sense in the text. The 15 types of relationships are: Solution-hood, Evidence, Justification, Motivation, Reason, Sequence, Enablement, Elaboration, Restatement, Condition, Circumstance, Cause, Concession, Background, and Thesis-Antithesis (Mann & Thompson, 1986).

By contrast, drawing on theories of the Prague School of Linguistics and the analysis of topical structure to describe coherence in texts, the Finnish linguist Lautamatti (1987) examined how topics repeat, shift, and return to earlier topics in discourse. They developed a model of topical sentence structure using topic-comment analysis to examine written text. One kind of system to achieve this topical structure for Lautamatti is sequential progression: "[s]he applies her analysis to written discourse to show that certain patterns of topical progression may be more readable than others (i.e., those texts which have fewer competing subtopics, fewer complex sequential progressions (A-B, B-C, C-D) are more readable)" (Grabe & Kaplan, 1998).

Finally, since our study specifically concerned narrative composition, we also paid attention to Labov's Narrative Model (1972) which provides a schema of the typical episodic elements which make up a coherent story: Abstract – Orientation – Complication – Evaluation – Result – Coda. This model has been widely used in narrative discourse studies ever since.

Method

This study attempts to identify and explain the source of salient error and performance features of the English compositions written by better and weaker Saudi English majors at the tertiary level. Although some attention was paid to the grammatical and lexical levels, the main focus was on four higher level discourse areas.

Participants and Context

The participants were Saudi educated, Arabic speaking, first year students in an English BA program in Saudi Arabia. This institution is typical of many in Saudi Arabia with respect to the English proficiency level of the students (in international terms, intermediate) and the focus of the English major program, which aims to further improve students' own English proficiency as well as educate students in areas such as linguistics, applied linguistics, and EFL.

There were initially 20 student participants who had been attending the Composition Two course in the English Department, whose institutionally stated objectives are to enable students to write: work-related memos and letters; directive, descriptive, narrative, coherence and informative paragraphs; and compositions on global themes. At the time of the study, 18 students showed up and agreed to participate. However, three participants did not write a sufficient amount of text (only two lines), and another was later excluded because they were found to have studied in the UK for three years, so did not match the target population. Hence the study included 14 participants with similar background experience in writing instruction in Saudi schools and English instruction in the preliminary year of the BA.

Although students would have had some experience writing short narrative compositions in English (and Arabic), to our knowledge they would have not received adequate instruction targeting narrative structure or cohesion and coherence. Rather, they had been instructed during their previous English and Arabic writing classes to write in the classical pattern of Introduction – Discussion – Conclusion. Based on the scores obtained for the compositions (see 4.4), seven participants were treated as relatively 'strong' and seven as 'weak' when reporting results.

The task

The topic of the 250-word narrative composition writing task was: "Describe a Wedding you were Invited to." This is similar, but not identical, to topics that participants would have written about before. The task falls within the previously stated objectives of the participants' Composition Two writing course with regard to topic familiarity, length of text, language, and allotted task completion time. Indeed, a similar task forms part of the final assessment of the writing course.

Procedure

Data was gathered by the researcher at the end of the Composition Two course in the English Department. The task was administered to students who consented to participate in a class out of regular class times. After the purpose of the research was explained, they were given the task title and instructions (in English), after which they had 30 minutes to complete the task under quasi-exam conditions: no dictionaries or collaboration between students were allowed. No incentive was offered to participate and the composition scores did not count towards the students' assessment for the course.

Data Analysis

The compositions were first independently scored by two experienced EFL markers using the local scoring system (awarding a global mark out of 20), which penalizes errors based on how much they affect the delivery of the intended message and not, for example, on how cohesive the units of the texts are. Although not an internationally recognized writing assessment instrument, we felt it was adequate for our purpose since it showed excellent inter-marker reliability (Pearson r=.954), and good concurrent validity, given that mean scores from the two markers correlated well with recognized indicators of writing quality, such as word length of composition (r=.633, p=.015), and number of different words used (r=.644, p=.013). Accordingly, the participants were categorized as relatively 'strong' or 'weak' writers depending on how they scored relative to the institutional pass mark of 14 (strong students M = 17.8; weak students M = 11). Next an analysis was undertaken of lower level errors and performance features conventionally examined in analyses of EFL written products (as reported in Table 1).

Finally, and most importantly, the compositions were analysed from four discourse perspectives. First, we identified what narrative patterns, in terms of Labov (1972), were mostly preferred by the participants and how the two groups differed in their patterns. Second, a detailed analysis of the participants uses of cohesive devices was carried out, following the scheme of Halliday and Hasan (1976), with attention to the appropriateness of the devices used. Additionally, for the different types of cohesion a difficulty index was calculated as a percentage based on the number of incorrect uses of a device divided by the total number of uses multiplied by 100. For example, if students used a type of cohesive device 80 times and made 20 mistakes, then the index of difficulty for the device is 20*100/80=25%. Thirdly, the coherence of the compositions was analysed in terms of conceptual relationships between propositions, in the style of Mann and Thompson (1986). Finally, inspired by Lautamatti (1987), the styles of progression system found were identified in relation to topic continuity.

Results

Basic Error and Performance Features

Although the strong writers scored above the institutional pass mark, their compositions nevertheless contained some misuse of grammatical and lexical features. By the same token, the compositions of the weak writers included some appropriately used grammatical and lexical features. Due to space limitations we will not elaborate on these in detail, but make the important point that this should alert us to the fact that weak participants do not totally lack writing ability. Indeed, even an example such as the following, from a weak participant, although containing serious lexical and grammatical/structural shortcomings, due in part to L1 interference, nevertheless exhibits elements of textuality, cohesiveness, and coherence at a higher level: *I am very hope for Ahmed's father spent many and food*. This sentence can possibly be construed as meaning "I really hope that Ahmed's father has spent a lot (on the wedding) and supplied food." This supplies additional justification for our aim in this study to focus on higher level discourse features contained within the text quite apart from the appropriateness of the specific sentential grammatical and lexical forms. Such lower level mistakes should not preclude analysis and evaluation of text at higher levels.

With respect to basic performance features, Table 1 shows a mixture of features where there were clear differences between strong and weak writers and others where there was no real difference. On average, strong students displayed higher total numbers of different words used and of sentences written than weak ones, and were on average not far short of writing the target number of words set for the task (250). This is an indication that these participants probably had both greater lexical and grammatical knowledge and were able to retrieve it more easily, allowing them to write more.

Nevertheless, the groups did not differ much in average word length, showing that neither group was employing more difficult words than the other (since on average harder and rarer words are longer). Similarly, average sentence length did not differ significantly, suggesting that neither group used more complex syntactic structures. Type-token ratios were also similar and quite low; this means that the repetition rate was high: both groups were using a small set of words repeatedly. This was probably partly due to the topic and the text genre, but equally because they return to the same ideas while narrating the events in a parallel manner. This may result from the rhetorical features of the Arabic writing style (Kaplan, 1966, 1972). Overall, then, the strong and weak writers do not differ on these measures of language complexity or sophistication.

Table 1
Participants' Scores on Five Basic Performance Measures, with Results of the Independent Samples T-test Comparing the Groups

	Group	Mean	SD	t	p
Total number of running words	weak	148.0	65.5	2.395	.034
	strong	221.1	47.3		
Total number of sentences	weak	15.4	2.9	2.834	.015
	strong	22.9	6.3		
Mean word length in letters	weak	3.9	0.18	1.204	.252
	strong	4.0	0.18		
Mean sentence length in words	weak	9.98	4.7	0.005	.996
	strong	9.99	2.3		
Type token ratio	weak	.523	.02	0.627	.542
	strong	.517	.01		
Total different word types	weak	77.3	33.4	2.398	.034
	strong	114.1	23.1		

Narrative Patterns

The participants used a wide variety of narrative patterns identified by Labov's Narrative Model (1972), even though they had been instructed during their previous English education and L1 (Arabic) instruction only in the classical pattern: Introduction, Discussion, and Conclusion. We found that all participants (strong and weak) exhibited basic knowledge of narrative composition structure by being able to implement the fundamental components of Labov's Model, i.e., Abstract – Orientation – Complication – Result. There were, however, notable differences between the pattern elements found in the texts of strong and weak writers.

Some strong participants had developed a systematic pattern by which they expanded on the core story components of Labov's Model to include additional components of Evaluation and Coda. Although they had not been taught such features, it is clear they had acquired enough skill and knowledge to develop their compositions in this way (e.g., strong Ss #4, 9, & 13). As an example, consider participant #4's composition which contained the following sequence:

> Abstract: Last week, I was invited to my friend's wedding, when I went there, I saw many things.
> Orientation: My friend is the groom and his tribe is not same my tribe. So he has a little bit different traditions of me….etc,

Complication: In fact, I saw a strange thing and it was interesting. There was big quantity of food for the dinner. When I entered the place I saw a beautiful place. When the women saw me, they took me…etc.
Evaluation: But I went to attend a friend's wedding to see that interesting place…etc.
Result: At the end I was very happy
Coda: I hope that happens to you in the future.

Such text organization reinforces coherence and strengthens textuality but this case is very rare among the participants' compositions, whether written by strong or weak writers. Few participant compositions displayed a truly logical sequence of episode development where events are introduced, exemplified, detailed, explained, and supported with general-specific instances, and where sub-topics and plots are developed, then elaborated upon with an evaluation and final resolution. Many (especially weak) participants exhibited recurring and overlapping components rather than a linear development (Kaplan, 1966), due to interference from their L1. One example of this feature is seen in the composition by a strong participant, Participant #3, where the Abstract overlaps with Orientation, Complication recurs three times, a second Orientation is included, and Evaluation is repeated three times after each Complication: "I am very hope for Ahmed because he did my advice but I saw bad thing in that wedding. I saw Ahmed's father spent extra money and food."

One may argue that it could be an advantage to have such a variety of components within the composition if the variety is based on the sequence of events and on the need to elaborate on sub-topics and themes individually. This variety is, however, not consistent in the sense that it does not occur in parallel with the individual sub-topic and theme development. Some parts of the one composition have consistent patterns internally, but others do not, which results in poor organization.

Cohesion

Many kinds of cohesion were observed in the participants' compositions, both strong and weak, though the strong students displayed greater ability to implement cohesive devices. The analysis further revealed that shortcomings in the form of misuse and absence of cohesive devices results mainly from the impact of L1 interference of Arabic rhetorical characteristics such as parallel structures and coordinate sentences, as previously found (Kaplan 1966, 1972). General lack of control over grammatical structure and poor vocabulary knowledge also contributed to weakening the cohesion of the participants' written texts.

Reference

There was considerable misuse of pronouns of reference in both groups' compositions, mainly due to L1 interference of Arabic grammar. Examples (1.1) – (1.3) are examples of strong participants' mistakes. Examples (2.1) – (2.5) are examples of weak participants' mistakes.

(1.1) Resumptive pronoun in an adjectival clause:
I sat down near my old friends Ali and Omar whom I haven't seen them since 1995.
(1.2) Direct personal pronoun in an indirect question:
I asked him why are you worry?
(1.3) Wrong choice between *other/the other/another*:
a. My father asked my brother if he invited another people.
b. Some of them comes from Qatar, the other came from Kuwait.

(2.1) Resumptive pronoun in an adjectival clause:
…suddenly found somebody who I not met him since…
(2.2) Direct personal pronoun in an indirect question:
They asked him how many chairs you will make a reserve.
(2.3) Wrong choice between *other/the other/another* forms

a. Some of them from the outside of the Kingdom and the other from Saudi Arabia.
b. My friend took me to the other place and showed me some dance.
c. …then, somebody go to the home, and another body no go home.

(2.4) Absent participant pronoun:
After one hour – give the dinner in body the eat…

(2.5) Unclear pronoun reference:
When I arrived to the hotel I met an old friend who was with me and the groom in the school and he also was inviting to the wedding.

There is considerable similarity between mistakes made by both categories of participants. However, the strong participants are distinguished by more appropriate sentence structure and vocabulary usage, which compensate in some measure for reference mistakes.

Substitution

Halliday and Hasan (1976) define substitution as a formal lexicogrammatical relation in which a form (word or words) is specified through the use of a grammatical signal which shows that it is to be recovered from what has gone before.

We found first that the participants (both strong and weak) used substitution, whether nominal, verbal, or clausal, at a minimal rate, and, if used, it was often inappropriate due to the fact that the participants had not acquired the necessary linguistic competence to implement such a device appropriately. Participants borrowed L1 written patterns because they lacked knowledge of the use of substitution as a cohesive device.

Second, participants (both strong and weak) often repeated the same lexical items instead of using substitution. Again this is a prominent L1 (Arabic) rhetorical feature. For example:

(3.1) Last week, my friend invited me to his wedding. I came to wedding and …
(3.2) Then, we went to the tent for eating dinner in fact the dinner was very good.
(3.3) I will describe the wedding. The wedding was in medium area…

The above weaknesses contributed to the production of poorly cohesive compositions. Overall, the participants used substitution in 32 instances appropriately versus 12 inappropriate instances in a total of 268 sentences contained in all the compositions. Substitution ranked third in the difficulty index based on errors with the cohesive devices.

Ellipsis

Similar to substitution, ellipsis is either nominal, verbal, or clausal. Halliday and Hasan (1976) define ellipsis as “substitution by zero”: "Where there is ellipsis, there is a presupposition in the structure, that there is something to be supplied, or understood" (p. 143). Like with other cohesive devices, the participants widely translated or transferred their L1 (Arabic) structures in their compositions (Connor, 1996). Consider an example of intended ellipsis but with an Arabic influenced structure (4.1). The writer tried to use an elliptical structure in the second clause but failed due to L1 interference. In written English we would expect overtly marked cohesion here *…and his brothers (were) too*. Similarly, in (4.2), we can clearly see transfer from Arabic structure in the second clause, but in this case it is successful as English works in the same way here, due to the use of the word *also*.

(4.1) I saw Ahmed's father very happy, and his brothers...
(4.2) Not only were the groom and bridegroom happy, but also the invited people.

The distinctive features of ellipsis used by both strong and weak participants included dropping the repeated verb, reversing sentence order and a tendency to avoid using ellipsis when it is necessary. The reason for not using this device may be due to the belief that ellipsis is not encouraged in written language as its use is more a feature of speech.

Overall, there were 25 instances of different kinds of ellipsis, 15 of which were appropriate and 10 were not. Implementing ellipsis as a cohesive device ranked first in the difficulty index with the least frequent number of correct occurrences.

Conjunction

The conjunctional connecters are the cohesive devices perhaps most commonly used by most speakers and writers alike. For Arabic speakers, conjunctions are a frequent rhetorical feature of oral as well as written language. Hedge (1988) states that "classifying these devices is made difficult by the fact that many have grammatical functions as well as logical ones" (p. 92). They link parts of the text grammatically and logically, clearly showing the natural flow of the text for the reader. Thus, many teachers have laid great emphasis on teaching conjunctions as a crucial element responsible for text cohesion (McCarthy, 1991). For this reason perhaps, to a large extent, the participants of this study used conjunctions satisfactorily. They produced 92 instances, where 72 were provided by the strong participants and 20 by the weak ones, with a total of 44 incorrect instances. This means conjunction is ranked fourth in the difficulty index.

Examples (5.1) – (5.4) show the following features: (a) Conjunctional markers are used to link short coordinate sentences. This is a well-known Arabic (L1) conjunctional feature. (b) The level of connectedness is fragile and superficial. (c) Conjunctions are sometimes used instead of referential or other relational ties, e.g., *and* instead of *to* or *who* – oral Arabic features (Example 5.3). (d) Conjunctions are placed in the second clause after a full-stop instead of a comma, an Arabic rhetorical written and oral feature (Example 5.2). And (e) the same conjunction is used to link clauses and phrases in the same sentence (Example 5.1).

(5.1) I came to the wedding and met the groom and bridegroom.
(5.2) The wedding was far away from the city. But I went to attend my friend's wedding and to see that interesting place.
(5.3) Then there were the singers and they make the party very interesting.
(5.4) I went to the groom and I gave him the present that I bought.

The conjunctional ties count shows that 50.5% of the markers used were additive, 38% temporal, 6% causal, and 5.5% adversative. The high rate of additive conjunctions is due to ease of use of the tie and the influence of L1 (Arabic) use of such markers in both oral and written discourse.

Lexical Cohesion

Two kinds of lexical cohesive devices are identified by Halliday and Hasan (1976): collocation and reiteration. While collocation refers to the restrictions on how words can be used together, especially in phrases, reiteration refers to the use of related lexical items instead of reference or substitution. The prominent feature of lexical cohesion in this study is the participants' failure to implement it successfully due either to a lack of vocabulary knowledge or L1 interference, but not actually failure to attempt to use such cohesion.

The following examples illustrate this point. In Example (6.1), the sort of parallel structure with repetition of the same predicate exhibits a familiar Arabic rhetorical lexical cohesive feature. It also constitutes a form of reiterative lexical cohesion found in English, though not commonly used. In our data there were many such pairs, evidencing a straightforward impact of L1. Arab speakers and writers tend to use this rhetorical structure in their speech and writing to impress and emphasize what they want to say. In Example (6.2), a lack of lexical cohesion arises from the use of a lexical item that makes no sense in the context, or one so wrongly spelt as to be unrecognizable, due to poor lexical knowledge. In Examples (6.3) and (6.4), the lack of lexical cohesion arises from faulty collocation. One *arrives* at a place, not a person, and one does not *make* a wedding in a hotel but rather holds it there. Such features weakened the lexical cohesion and distanced it from native speaker norms.

However, in this area the performance of most of the participants is similar and this form of cohesion scored the lowest on our difficulty index.

(6.1) The wedding was beautiful. The bride was beautiful.
(6.2) My friend took me to the guest to talk reciprocal.
(6.3) When we arrived to the groom, we congratulated him and set together in the same table.
(6.4) I decided to make my wedding in the same hotel.

Coherence

Overt cohesive markers

Both high and low level participants often used linguistic signals (i.e., cohesive devices: 5.3) to relate propositions although such relationships are believed to be ultimately independent of lexical signals and need not always be overtly marked. The weak participants relied very much on such signals to mark thematic relations between each pair of clauses; the strong participants, on the other hand, displayed more ability to compose propositional relationships with a minimal use of signals.

Examples (7.1) – (7.3) are from strong participants. Example (7.1) conveys a causal relationship marked by the cohesive conjunction *so*. The sentences could however have been written without *so* because in this case one can easily pragmatically infer the causal relationship from the content of the propositions even if *so* is absent. In Example (7.2), a reason relationship is marked by the cohesive conjunction *because* again the relationship could be recovered even if not marked linguistically. In (7.3), again the relation may be construed as reason. However, there is no need for an overt cohesive marker and none is used.

(7.1) My friend is the groom and his tribe is not same my tribe. So he has a little bit different traditions of me.
(7.2) I was very happy because Mohammed was groom.
(7.3) I was invited to a wedding party form my friend. I wore a white thobe and a white ghutra.

Examples (7.4) – (7.5) are from weak participants. In Example (7.4), the reason relationship is present but not marked while in the second (7.5) it is marked with *because*. Furthermore, in Example 7.5, there is a concession relation, where a clause is introduced overtly with *but*.

(7.4) When he saw her he was surprised she was wearing a beautiful dress and she was holding a rose.
(7.5) I am very hope for Ahmed because he did my advice but I saw bad thing in that wedding. I saw Ahmed's father spent extra money and food.

The technique adopted by the participants' results mainly from (a) L1 interference, since it is an Arabic rhetorical feature to mark a relation with an overt cohesive signal and (b) lack of sufficient linguistic knowledge to form sentences which are grammatically independent of others yet related propositionally to previous or successive ones.

Elaboration

Elaboration was one of the most prominent propositional relations employed by the participants. Examples (8.1) and (8.2) are from strong participants. Example (8.1) suggests the whole-part relationship and (8.2) involves set-member relationship.

(8.1) I saw a strange thing and it was interesting. There were a big quantity of food for the dinner.
(8.2) There were many drinks. There were coffee, tea and all the refreshments.

Examples (8.3) and (8.4) are from weak participants. In (8.3), the second sentence is an instance proposition. In (8.4), the second sentence could be an attributes proposition.

(8.3) My brother looks like a new man in wedding day. He bought a new watch for the night.
(8.4) There is many people come to his wedding. I saw people came from Jeddah and from Abha and Makkah.

One explanation for the extensive use of elaborative propositional relations by both strong and weak participants equally could be the nature of narrative discourse, imposed by the set topic. It could also be attributed to the participants' tendency to provide evidence and examples for the fact or events they include in their writing. Whatever the reason is, elaboration in itself re-enforces coherence between the units of the text.

Temporal sequence

An outstanding introductory signal of relationship used by the weak participants is cohesive conjunctions indicating temporal sequence (Examples 9.1 – 9.4).

(9.1) While I was there I met a lot of people some of them from the outside of the Kingdome and the other from Saudi Arabia.
(9.2) When I was there I saw many people who were invited like me.
(9.3) Before the party of the wedding finished, my friend who is groom ask me about what did he do with his bridegroom…
(9.4) When we arrived to the groom we congratulated him and sit together in the same table…

These examples all display a simple temporal relationship between each pair of clauses. The reason for the use of such clauses are the effects of L1 Arabic rhetoric where emphasis of temporal mode is a well-known spoken and written feature. It is believed that speech is more convincing when such signals are used. Such signals are however generally used more functionally in English, i.e., writers would not use them if there was no real need (see for instance Bern,1988).

In our study, such signals are not used as frequently by the strong participants. The strong participants employ clausal relations using more appropriate ties when there is an actual need (Examples 10.1 – 10.4).

(10.1) There was a quantity of food. This is because the groom invited a lot of people.
(10.2) I was invited to a wedding party form my friend. I wore a white thobe and a white ghutra.
(10.3) Once, I was invited to a wedding party, I was really very pleased.
(10.4) My uncle had brought a singer. The singer sings good. He had sang many wonderful songs.

The weak participants are thus distinct from the strong ones due to less varied relatedness among sentences.

Topic and Progression System (Topic Continuity)

Finally, we consider how each group of participants developed their topics across paragraphs (following Lautamatti, 1987). A remarkable feature creating coherence in the strong participants' compositions is the maintenance of the progression system by starting a new topic with a comment associated with the previous topic (Example 11.1).

(11.1) The dinner was served at 9:30. its was very tasty and well-cooked food. After the dinner, a group of men danced and sang for three hours. This dance is call Al-Ardha and it has two sides. In the first side, there were about 15 men who were singing. In the second side, there were about 20 men who were carrying swords on their shoulders.

Examples from weak participants such as in (11.2) show a different progression system, which includes an extended parallel progression where the same sentence/clause topic (here *I*) is repeated.

(11.2) I saw many good things in that wedding. I saw the groom and his family. I saw my old friends and I talk to him. I remember what we did in the past with him. I ate a good and big dinner. After the dinner I make a good argument with my best friend to help my friend Ahmed is very popular man. There is many people come from Jeddah, and Abha and Makka.

Discussion and Conclusion

Summary of Findings

RQ1. What are the prominent non-nativelike features of the written English narrative compositions of Saudi English major first year university students (of intermediate proficiency)? Especially what do we find at the higher levels of rhetorical structure/text organization? The examples cited throughout exhibit many lower level grammatical and lexical problems, which it was beyond the scope of this article to describe fully. At the discourse level in general our participants exhibited the core elements of narrative structure despite not having been explicitly taught it, but neither strong nor weak writers consistently achieved an elaborated and properly sequenced episodic development. Among the cohesive devices, conjunctions tended to be employed in a more native-like way than ellipsis, with reference and substitution in between. Coherence was characterized by tending to be marked cohesively even where not necessary, and by reliance on elaboration. We conclude therefore that first year Arab EFL BA students, regardless of their scores on composition tests, are not necessarily capable of producing cohesive and coherent texts that meet well-known global writing standards, based on analyses using current discourse analysis frameworks. To an extent this then supports the view of Kaplan (1966) that transfer of rhetorical patterns occurs regardless of proficiency level.

RQ2. How much and in what specific areas do strong and weak writers differ? The general finding from the compositions (Table 2) is that there are distinct differences as well as similarities between the performances of the strong and weak participants, thus supporting Witte and Faigley (1981) who found that good essays included more cohesive devices, but going against Zhang (2000) who found that there was no difference in the appropriacy of the use of cohesion and coherence devices between strong and weak compositions.

Table 2
Summary of Main Group Differences

Feature	Strong writers	Weak writers
Composition score	Higher score with low variation	Lower score with higher variation
Fluency: Numbers of running words, sentences, and different word types used	Higher	Lower
Lexical and grammatical complexity: Mean word length, mean sentence length, type-token ratio	Similar	Similar
Grammatical correctness	Higher	Lower
Narrative structure	More sophisticated but not necessarily properly sequenced	Basic
Cohesive devices	More fully and appropriately used	Less appropriately used
Coherence relations	More varied and inexplicit	Less varied and more overt
Topic progression	More sequential (A-B, B-C)	Less sequential and more parallel with repetition (A-B, A-C)

RQ3. Is the reason for non-native-like performance by these writers primarily transfer from L1 Arabic or other factors? There was considerable L1 interference of Arabic linguistic and rhetorical features in the compositions produced by participants in this study. This contradicts Ismail (2010), but is consistent with the majority of earlier work such as that of Connor (1996), Kaplan (1966, 1972), Ostler (1987), and many others, who highlighted L1 rhetorical transfer. Nevertheless, we believe that there is a range of other factors which contributed to what we observed when analyzing the texts, including some mentioned by Ismail (2010).

For example, instruction has probably played a role in the greater successful use of conjunctions than other cohesive devices, due to greater focus on this area in prior teaching, and in the weaker performance on aspects of cohesion and coherence that are not taught. This is due not only to the current program in our participants' first year but also their previous educational background, which is largely determined by the prescribed English textbooks used in school and the Preliminary (pre-BA) year. In our context, teachers are required to follow prescribed textbooks and are given little scope to introduce additional topics into their teaching. Due to differences, and sometimes conflicts, between English speaking Western cultural concepts and those of the students' local culture, the instructional materials are carefully selected and mostly edited to avoid conflict with the native culture. This has contributed to a general lack of English authentic materials in Saudi EFL programs, which limits students' exposure to L2 discourse patterns.

The quality of linguistic performance in this writing task may also have reflected how far students have accessed, or had access to, English books, media, and international speakers of English in addition to their traditional instruction. Students have very little opportunity for such involvement unless they choose to use the Internet for this purpose. Language instruction is provided within a highly structured and traditional paradigm largely cut off from the wider English speaking community outside the country.

Our Arab EFL students also face constraints deriving from their generally limited linguistic proficiency in the L2. Limited knowledge of fundamental vocabulary and grammatical structures constrains L2 writers' performance at higher discourse levels, especially in the area of explicit cohesive devices.

Pedagogical Implications

In light of our findings, we present some recommendations to relevant stakeholders.

The participants in this study displayed limited linguistic and rhetorical competence at many levels. Consequently, the written work of graduates often falls below the standard expected by the administration. The features we targeted should be understood by EFL curriculum designers, textbook writers, and teachers in Arabic contexts, so they can help students improve in order for them to achieve the expected standard, and as a result, to have greater success in majors delivered using English, in future careers, or study abroad programs.

One way of making students more aware is to implement language awareness tasks. In effect, EFL students can be provided with mini-research tasks on L1 native texts and their own compositions, which will alert them to the rhetorical features of those texts and so to the differences between L1 discourse and their own practices influenced by L1 interference.

Along with these recommendations, it would be beneficial if the EFL instructional syllabus contained more authentic materials to provide greater exposure to L1 written texts. Furthermore, EFL writing instructional practices should be adjusted to allow for teacher innovation and the introduction of supplementary tasks and materials as required.

Limitations and Future Research

Our sample was quite small, and had this study been conducted in locations additional to the available setting, broader population representation would have been achievable. The findings also need to be replicated with female learners. Furthermore, discourse analysis is such a broad topic, with many available theories and methods of analysis, that there are certainly other issues and further variables that could be explored. With more time, it would also be valuable to interview participants to ascertain in more detail their wider exposure to English outside the confines of the classroom to better illuminate the factors underlying their performance. Nevertheless, we feel that our study paves the way for future research and our narrative discourse analysis procedures are replicable in any other wider-scale study that may involve more data.

References

Alissa, M. (2013). EFL writing composition in the preparatory year: Cohesion and writing quality. *Journal of Language Forum (Bahrain University), 3*, 109-131.

Bern, D. (1988). *Teaching writing skills*. London, UK: Longman.

Connor, U. (1996). *Contrastive rhetoric*. Cambridge, UK: Cambridge University Press.

Currie, G. (1990). *The nature of fiction*. Cambridge: Cambridge University Press.

Grabe, W., & Kaplan, R. B. (1998). *Theory and practice of writing: An applied linguistic perspective*. New York: Longman.

Halliday, M. (1994). *An introduction to functional grammar*. London: Edward Arnold.

Halliday, M., & Hasan, R. (1976). *Cohesion in English*. Harlow, UK: Longman.

Halliday, M., & Hasan, R. (1989). *Language, context and text: Aspects of language in a social-semiotic perspective* (2nd ed.). Oxford: Oxford University Press.

Hedge, T. (1988). *Writing* (2nd ed.). Oxford, UK: Oxford University Press.

Ismail, S. (2010). *Arabic and English persuasive writing of Arabs from a contrastive rhetoric perspective* (Unpublished doctoral dissertation). Indiana University of Pennsylvania, U.S.A.

Johnson, D., & Roen, D. (Eds.). (1989). *Richness in writing: Empowering ESL students*. London and New York: Longman.

Kaplan, R. B. (1966). Cultural thought patterns in intercultural education. *Journal of Language Learning, 16*(1), 1-20.

Kaplan, R. B. (1972). *The anatomy of rhetoric: Prolegomena to a functional theory of rhetoric*. Philadelphia: Center for Curriculum Development (distributed by Heinle and Heinle, New York).

Kroll, B. (Ed.). (1990). *Second language writing*. Cambridge, England and New York: Cambridge University Press.

Labov, W. (1972). The transformation of experience in narrative syntax. In W. Labov (Ed.), *Language in the inner city: Studies in Black English vernacular* (pp. 354-396). Philadelphia: University of Washington Press.

Lautamatti, L. (1987). Observations on the development of the topic of simplified discourse. In U. Connor & R.B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text* (pp. 87-114). Reading, MA: Addison-Wesley.

Mann, W.C., & Thompson, S. A. (1986). Relational propositions in discourse. *Journal of Discourse Processes, 9*, 57-90.

Mauranen, A. (1993). *Cultural differences in academic rhetoric: A textlinguistic study*. Frankfurt: Peter Lang.

McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge: Cambridge University Press.

Ostler, S. (1987). English in parallel: A comparison of English and Arabic prose. In U. Connor & R.B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text* (pp. 169-185). Reading, MA: Addison-Wesley.

Raimes, A. (1991). Out of the woods: Emerging traditions in the teaching of writing. *Journal of TESOL Quarterly, 25*, 407-30.

Rammel, S. (2009). *Rhetorical issues in the writing samples of Palestinian learners of English*. Retrieved from http://www.academia.edu/1474953/Rhetorical_Issues_in_Samples_of_Arab_Students.

Reid, J. (1993). *Teaching ESL writing*. Englewood Cliffs, NJ: Regents Prentice Hall.

Saadeddin, M. (1989). Text development and Arabic-English negative interference. *Journal of Applied Linguistics. 10*(1), 36-51.

Stapa, S.H., & Irtaimeh, M. (2012). Rhetorical awareness and Jordanian EFL students' expository writing. *European Scientific Journal, 8*(26).

Witte, S., & Faigley, L. (1981). Coherence, cohesion and writing quality. *College Composition and Communication, 32*, 189-204.

Zhang, M. (2000). Cohesive features in the expository writing of undergraduates in two Chinese universities. *RELC Journal, 32*(2), 61-93.

Schematic Structure and Linguistic Characteristics of Discussion Sections in Applied Linguistics Research Articles

Attapol Khamkhien*
Kasetsart University, Thailand

Abstract

Research articles written in English are increasingly used in advanced education as a means to disseminate and share knowledge among scholars. Inspired by the analytical framework of Swales (1990, 2004), this study analyses a set of Discussion section texts in applied linguistics research articles written in English with regard to their rhetorical patterns and linguistic features. Ten research articles were systematically selected from the journals indexed in the Thai-Journal Citation Index (TCI) database, and were analysed using move analysis. The analysis reveals that the Discussion sections consist of three move types, representing distinct communicative purposes. Linguistic and lexico-grammatical features salient in individual moves and steps are highlighted. The findings of the study help novice scholars in general and Thai graduates in particular, to better understand what to include when writing a research article. The findings of this study are pedagogically beneficial in providing Thai graduate students with the skills required to disseminate their knowledge, useful for preparing them with the general expectations of the discipline, and facilitating the task of writing and reading research articles.

Keywords: Discussion section, research articles, move analysis, applied linguistics, Thai-Journal Citation Index (TCI) database.

Introduction

As the English language has achieved the status of an international language, the ability to use it effectively has become essential. Within any academic discipline worldwide, English is used in diverse forms to communicate, particularly in higher education (Coleman, 2006; Crystal, 2004). Among the numerous means of communication, it is commonly agreed that the role of research articles (RAs) written in English is becoming increasingly important. RAs are used to share new discoveries and knowledge in the academic community. Therefore, to survive academically, scholars and researchers must be actively involved in the writing of RAs and other academic publications.

In Thailand, being able to write for scholarly publications is becoming ever more important. Several strategic plans initiated by the Thai government have been used to encourage Thai scholars and novice writers to publish their research findings in international journals, including the establishment of the National Research Fund in 1992, the Royal Golden Jubilee Program in 1996, and the Office for National Education Standards Quality Assurance in 2004 (Chalapati, 2007). However, according to Sinlarat (2004), these strategies seem to have been unsuccessful as academics in Thailand still only publish a small number of research articles in international journals. With regard to Thai graduate students, having their research work published in prestigious and peer-reviewed journals seems to be crucial as it is a graduation requirement (Svasti & Asavisanu, 2007). Many Thai universities require candidates wishing to graduate with a master's or doctoral degree to have their work published in a refereed journal. According to Kanoksilapatham (2011, p. 56), writing an RA is a daunting

* *Tel. 66 3428-1105-7; Email: faasapk@ku.ac.th; Address: Department of English, Faculty of Liberal Arts and Science, Kasetsart University, Nakhon Pathom 73140, Thailand.*

task, not only for non-native speakers, but also for native speakers of English. Specifically, Thai graduate students seem to have some difficulties in having their work published in English. That is, the schematic structure of RAs written in English and the style of writing required, including language proficiency and the language itself, are particular obstacles that Thai graduate students encounter (Kanoksilapatham, 2007a). Hence, it is inevitable that the ability to write an RA in English that conforms to a set of standardized expectations is pivotal for novice authors in general, and for Thai graduate students handling academic writing tasks and publications in particular.

In this regard, a review of genre-based studies (e.g., Basturkmen, 2012; Swales, 1990, 2004; Peacock, 2011) focusing on 'move analysis' as articulated by Swales (1990, 2004) could reveal the hierarchical schematic structures of texts. These studies have provided invaluable insights into the general ways in which RAs are organised, and what should be included in each section. These studies, however, focused on the different conventional sections of RAs, namely, Introduction, Methods, Results, and Discussion (commonly known as the prototypical IMRD model) in various disciplines. Among these different sections, scholars investigating the Discussion section in several academic disciplines agree that this particular section is a significant part of RAs for a number of reasons and has a number of functions (Basturkmen, 2012; Kanoksilapatham, 2012; Yang & Allison, 2003). It is the section in which the researchers, based on the results, not only report the findings of the study being presented, but also explain, interpret, and compare results with those of previous studies. In this regard, a genre-based approach has been used to investigate the Discussion section in RAs from diverse disciplines; for example, medical science (Nwogu, 1997), engineering (Kanoksilapatham, 2012), and applied linguistics (Yang & Allison, 2003). The findings of these studies have shown distinctive patterns and moves, and revealed that disciplinary variation is discernable, leading to the incomplete detailed description of this particular section in RAs. Therefore, more studies are needed to determine if the structural pattern of this section varies according to the academic discipline in question.

The present study is primarily concerned with the schematic structure and linguistic features of the Discussion section in applied linguistics RAs. This work aims to provide scaffolding support and encouragement for Thai novice writers and graduate students in enhancing their opportunities for international scholarly publication, especially in journals indexed in the Thai-Journal Citation Index (TCI) database; moves and steps constituting the generic structure of the Discussion section will be analysed. Moreover, the extent to which linguistic choices can be examined in relation to rhetorical moves and constituent steps found using Swales' schematic framework (1990, 2004) will be investigated. This paper also examines linguistic features commonly used in this particular section. Pedagogically, the findings obtained will help facilitate the teaching/training of graduate students and novice scholars to construct the Discussion section in a manner conforming to the expectations of their respective academic communities, both rhetorically and linguistically.

Move Analysis

Move analysis, one of the genre-based approaches, is suggested to identify the structure of RAs. Swales (1990, 2004) points out that a text representing a genre consists of a discoursal segment called a 'move' sequenced in a particular order, forming a pattern. Each move possibly consists of sub-units called 'steps' by Swales (1990) or 'strategies' by Bhatia (1993). A move or step performs its own particular communicative function, recognised by a set of linguistic features. In other words, a move captures the function and purpose of a segment of text at a more general level, whereas a step presents more specifically the rhetorical means of realising the function of the move. Swales proposed a structural organisation, or move structure, for RA Introductions called "Create a Research Space" (CARS) in 1990 (revised in 2004), as shown in Figure 1.

Move 1	**Establishing a territory*** via**
	Topic generalisations of increasing specificity
Move 2	**Establishing a niche*** via**
Step 1A	Indicating a gap or
Step 1B	Adding to what is known
Step 2	Presenting positive justification*
Move 3	**Occupying the niche**
Step 1	Announcing present research descriptively and/or purposively (obligatory)
Step 2	Presenting RQs or hypotheses*
Step 3	Definitional clarifications*
Step 4	Summarising methods*
Step 5	Announcing principal outcomes**
Step 6	Stating the value of the present research**
Step 7	Outlining the structure of the paper**
	* *Optional and less fixed in order*
	** *Probable in some academic disciplines*
	*** *Possible cyclical patterning of moves particularly in longer Introductions*

Possible recycling of increasing specific topics

Figure 1. Swales' 2004 revised model for RA Introductions

Swales' analytical model was created to facilitate the task of writing RAs, especially for those whose first language is not English. In this regard, the model, especially moves and steps identified based on their functions, was criticised by scholars for its subjectivity. Conducting a move analysis also lacks explicit rules for decisions on move boundaries, leading to questions of the reliability and empirical validity of the analysis (Paltridge, 1994). However, as suggested by Dudley-Evans (1994), decisions about the classification of move patterns was made on the basis of linguistic evidence, comprehension of the text, and understanding of the academic discourse community. Furthermore, according to Swales (2004), the model proposed is not meant to prescribe, but is used to provide guidelines with regard to how ideas are marshalled when writing an Introduction. The model thus offers the schema and structural patterns of what informational elements should be included in order to meet the expectations of a particular discourse community.

Although criticism concerning the usefulness and validity of the CARS model has been raised, the model has been subsequently validated in a number of RA Introductions across several disciplines (e.g., Samraj, 2002 in wildlife behaviour and biology conservation; Laurence, 1999 in computer science; Kanoksilpatham, 2011 in civil engineering; Ozturk, 2007 in second language acquisition and second language writing). Furthermore, these RAs show the same schematic pattern consisting of the three major moves suggested, and disciplinary variation in the structural pattern was also observed. As an extension, the model was also used to analyse other internal sections of RAs in academic disciplines, e.g., Brett (1994) in the Results section of sociology articles; Peacock (2011) in the Methods sections in RAs from eight disciplines; Lim (2006) in the Method sections of management RAs; Yang and Allison (2003) in the Results and Discussion sections of applied linguistics RAs; and Nwogu (1997) in the Discussion sections of medical articles. The findings from these studies have demonstrated that each section of RAs has its own structural organisation. Additionally, these studies have helped us recognise that rhetorical organisation of texts is helpful, providing the schema for what information elements or moves are presented.

Moreover, even though each section seems to be written with similar rhetorical structure, genre-based studies across academic disciplines reveal that disciplinary variation is discernible.

By extension, Swales' (1990, 2004) analytical approach has also been applied to explore other professional genres, including application essays to medical and dental schools (Ding, 2007), research grant proposals (Conner & Upton, 2004), movie reviews (Pang, 2002), legal discourse (Bhatia, 1993), and replies to customer inquiries (van Mulken & van der Meer, 2005). Corresponding to Swales' (1990, 2004) move-based analysis, the findings of these studies suggest that knowing the rhetorical organisation of a text can be helpful, as they provide the schema for what informational elements or moves should be included and presented in each section of an RA. In short, these previous move-based studies of both academic and professional discourse provide a basic template for novice writers to disseminate their research results. The description and structural patterns of how each section of an RA is written allow writers to conform to disciplinary conventions and enhance their publication opportunities.

Knowledge and Skills Required for Publication

According to Flowerdew (2001), writing a research article in English can be a relatively daunting and challenging task, both for native and non-native speakers of English. This is because, in order to successfully write an RA, reading and writing skills are essential. That is, writers need to have read widely and have a clear understanding of published RAs in order to become familiarised with the state of the art of a field and to appreciate the expertise and contributions of individual scholars. Likewise, writing skills are essential to ensure that writers can efficiently and successfully express themselves academically. Cook (1990) and McCarthy (1991) suggest that, to be able to understand and produce research articles, writers need to make use of both macro and micro skills. The macro knowledge of a research article genre starts with the knowledge of structural organisation of the discourse units, before moving onto lexical and grammatical features used in a text. In contrast, the micro knowledge of the research article concerns the smallest units of discourse such as grammatical items, before moving onto more general features such as sentences, discourse units, and structural organisation of the discourse units. In other words, to write effectively, writers need to have knowledge of genre and linguistic characteristics, including lexico-grammatical features, rhetorical organisation, communicative functions, and content (Hyland, 2004).

These macro and micro language components have different, but complementary characteristics. Previous studies (e.g., Samraj, 2002; Kanoksilapatham, 2007a, 2011) demonstrated that each section of an RA, including Introduction, Methods, Results, and Discussion, had its own structural organisation, and is probably discernible with regard to disciplinary variation. Furthermore, patterns of lexical and grammatical features frequently used in research articles also tend to vary according to the writer's English proficiency and writing experience. Macro and micro levels of linguistic characteristics might also depend on the field and the expectations of the discourse community to which they will be addressed. Therefore, in order to assist Thai graduate students with scholarly publication, as suggested by Biber, Conner, Upton, and Kanoksilapatham (2007), authors need to improve their awareness of the kind of information usually presented, the sequence in which the information is organised, and what linguistic features are typically used to express each information element.

Methodology

This section consists of two main parts. Part one describes how the dataset of applied linguistic research articles was systematically compiled. From these original datasets, the datasets representing the discussion section were created. Part two describes how move analysis of the dataset in this study was conducted.

Dataset Compilation

The Thai-Journal Citation Index (TCI) Centre has established a database of journals published in Thailand, together with their citations online and in real-time, to encourage Thai scholars to publish their works and to ensure that the quality of the journals is on par with international standards. Ten RAs in the field of applied linguistics from three journals indexed in the TCI database were systematically selected for this study. The selection of these ten articles was controlled based on four considerations: the discipline of applied linguistics, type of RAs (experimental RAs), the internal sections of articles (IMRD), and availability of materials. Moreover, to minimise the researcher's subjectivity of individuals and to increase the validity of the analysis, the more objective criterion of the impact factor of the journals provided by the TCI database was consulted. At this juncture, it should be noted that the use of impact factor, though pervasively used in the sciences, has been challenged, particularly in the humanities and social sciences. The three selected journals were published in 2013. However, given that the focus of this study is on how the Discussion section is constructed, only the stand-alone Discussions of these articles were analysed. These three journals are referred to in this paper as HSNU (*Humanity and Social Sciences Naresuan University*), MP (*Manutsat Paritat: Journal of Humanities*), and HS (*Humanities and Social Sciences Journal*). These journals were approved to be included in the Tier one on humanities and social sciences by the TCI database. In addition, these journals are now on ASEAN Citation (ACI), representing the quality and reliability of the journals in particular fields. It should be noted, that since the statement of purpose sets the audience as the Thai academic community, the exclusion of the international weight of these publications was deemed acceptable. Moreover, even though the ten Discussion sections taken may not be representative of the given disciplines, to a certain extent, the analysis of the limited size of the dataset and linguistic features prevalent in texts can be considered useful. They serve as a preliminary indication of some trends displayed in RA discussion writing, and provide evidence of the structural organisation in applied linguistics RA discussions, and how linguistic features are used to accomplish their communicative functions.

Data Analysis

The dataset of research articles was analysed following the general steps of conducting a move analysis outlined by Biber et al. (2007). The objectives of this study are twofold: 1) to identify the rhetorical patterns of the selected applied linguistics research article Discussions based on a typical sequence of moves and steps, and 2) to identify linguistic features associated with the moves and steps identified. To address the first objective, the frequency of occurrence of each move and step identified in each Discussion section was recorded. Based on the most frequent sequence found, a typical structural organisation of the Discussions was formulated. To address the second objective, the alignment of individual move instances allowed us to identify a set of linguistic features commonly associated with each particular move and step in the Discussion sections.

Results

Working through the typical move analysis process proposed by Biber et al. (2007) and Swales (1990, 2004), the analysis enables us to identify three major moves which perform distinctive communicative functions. These three moves are presented in this paper in the order usually found in the dataset. Some identified moves contain constituent steps which are infrequently found, whereas some steps are frequently present. Table 1 summarises the move types and their constituent steps found to occur in this study.

In addition to identifying rhetorical moves based upon Biber et al. (2007), the move analysis of the Discussion section makes it possible to describe the typical linguistic characteristics of each move type identified above. The following sections present common features and linguistic characteristics of moves in the Discussions represented by each of the three moves in detail. To facilitate a better understanding, the description will be illustrated together with excerpts taken directly from the ten research articles, representing the function of each move and step. It should be noted that identification numbers for each sentence (S1, S2) have been added for ease of reference; second, citations used in the texts were replaced by *(Ref)* for space-saving purposes; third, all tables and headings were not included in the analysis; fourth, abbreviations for the names of selected journals were used; and lastly, lexical items and linguistic features which seem to particularly indicate communicative functions are highlighted in boldface.

Table 1

Proposed Model for Applied Linguistics Discussions

Move types	Number of articles
Move 1: Reviewing the present study	8
Move 2: Analysing and extending from results	10
Step 1: Reporting selected results	10
Step 2: Comparing results	10
Step 3: Explaining results	10
Step 4: Presenting arguments/positions	10
Move 3: Evaluating the study	10
Step 1: Stating value of the study	6
Step 2: Offering pedagogic implications	10
Step 3: Indicating limitations	8
Step 4: Suggesting further research	8

Move 1: Reviewing the present study

Move 1, if used, is found at the beginning of the Discussion section. This move serves as a reminder to potential readers of background information related to the study being reported, including the purpose(s) and major methodological features of the study, which are already established in the Methods section. Move 1 was found in eight research articles of the corpus. The following are examples taken from the beginning of the Discussion section (S1) that represent Move 1:

[1] (S1) *The first research question* ***was aimed to find*** *the extent to which Thai English teachers at* RMUTL *were of their use of metacognitive and cognitive reading strategies to understand English academic materials.*

JHNU 2

[2] (S1) *The present study* ***was designed to survey*** *low proficient students' attitudes towards language learning and casual attributions for their success or failure.*

HS 1

[3] (S1) *In this study, I particularly* ***focused on*** *'micro-interactions' (Ref) in that I* ***explored*** *the perspectives of FLTAs who were serving at a medium-sized state university in the Midwestern United States where particular norms, background, environment, and characteristics could inevitably affect students' behaviors in the classroom.* (S2) *In addition, I* ***explored*** *'particulars of individuals' lives' at a particular time (Ref) since my study* ***was conducted with*** *only students from three different foreign language classes in Fall 2009, i.e. Thai, Indonesian and Tagalog.*

HS 2

The excerpts above illustrate how readers are provided with sufficient background information to follow the discussion of the particular topic. This is accomplished through restating research questions and research purposes (*aimed to find*) as in [1], and methodological procedures (*designed to survey, focused on, explored, conducted with*) as in [2] and [3] before presenting the results, and commenting upon them. In addition, the use of verbs in past forms (*was aimed, was designed*, *explored, conducted*) indicates purpose(s) and methodology of the research studies being reported. With these strategies of providing readers with an overview of the studies, readers do not need to revisit the Introduction and Methods sections.

The present study found that Move 1 (Reviewing the present study) has no constituent steps, but recaps the objectives, research questions, and/or main procedures of the study. This finding confirms '*Move 1 Background Information*' found by Yang and Allison (2003) examining the results and conclusion sections of research articles in applied linguistics. However, it is not consistent with Nwogu's (1997) study focusing on the rhetorical organisation of medical research articles written in English. That study indicated that the Discussion section, '*Move 9 Highlighting overall research outcome*' only substantiated the success of the research objectives, but not objectives and methodological procedures.

Move 2: Analysing and extending from results

In general, Move 2 may be viewed as a reiteration of the summary of the results that the authors have reported in the Results section. Nevertheless, Move 2 goes beyond reporting the results, as its primary function is to strengthen the results of the study together with commenting on them and comparing them with previous studies. This move can be realised by one or more of these four steps. Step 1: Reporting selected findings, Step 2: Comparing results, Step 3: Explaining results, and Step 4: Presenting arguments and positions. All four of these are represented respectively as follows:

Move 2 Step 1: Reporting selected findings

Move 2 Step 1 is deemed the most central of this move, providing a basis for the other steps of Move 2. This move/step is crucial so that a finding can be introduced in order for it to be subsequently discussed in the Discussion section. Move 2 Step 1 was used regularly in all of the ten research articles chosen. A number of instances showing this move/step are illustrated below:

[4] (S4) *More interestingly, with regards to* ***Table 3 and Table 4****, the top ten strategies which were most frequently used* ***included*** *6 cognitive strategies and four metacognitive strategies*.

JHNU 2

[5] (S9) *They further* ***indicated*** *that participating in peer feedback activity helped them become more active and responsible for their own learning*.

MP 1

[6] (S8) *In addition, they were exposed to new ideas, different choices of words, and writing styles.* (S9) *They further* ***indicated*** *that participating in peer feedback activity helped them become more active and responsible for their own learning.*

MP 1

As can be seen above, to indicate this move/step, some linguistic signals associated with numerical values derived from statistical analysis are used. In instances [5] and [6], the reporting verb (*indicate*) is employed to report particular findings. (The reporting verbs could have been used in either the simple past or present tense in this move.) In addition, to aid the reader, referring to information shown in tables (*Table 3 and Table 4*) as in [4] is a feature of Move 2, Step 1.

Move 2 Step 2: Comparing results

Step 2 of Move 2 is to highlight the results being presented by comparing them with those of previous studies. Noticeably, reporting the particular results of the reported study in Move 2 Step 1 usually co-occurred with Step 2 of Move 2 (Comparing results), that is, the findings being presented are likely to be commented upon. Instances of how the authors compare the findings of the study being presented to the previous literature are presented below.

[7] (S6) ***This corroborates with (Ref)*** *who argued that if teachers do not have a thorough understanding of CLT, they can hardly develop practices appropriate to their context and thus they easily return to traditional teaching.*

JHNU 1

[8] (S23) ***This finding is contradictory to the studies*** *of* ***(Ref)***, ***(Ref)*** *and* ***(Ref)*** *which found that L2 students who came from a teacher-fronted classroom did not seem to welcome peer feedback.*

MP 1

[9] (S13) ***This finding is in line with (Ref)****'s study (2008) in which students who enrolled in fundamental English courses have a positive attitude towards English and exhibit positive motivation in English learning.*

HS 1

From the above, it is clear that this move/step can be identified by the use of references in order to compare and contrast the findings of the reported study with previous studies. Citations from previous studies are generally provided in order to demonstrate agreement and/or disagreement with the results *(Ref)*. Some linguistic features used to realise Move 2 Step 2 include the verb '*be*' and adjectives of comparison (e.g., *is consistent with, is contradictory to*, and *is in line with*). The use of the simple present tense (*corroborates, is*) is also found to be prominent in comparison and contrast with previous research.

Move 2 Step 3: Explaining results

The purpose of Move 2 Step 3 is to inform the reader of the possible reasons for explaining the results discussed in the Discussion section. This move/step provides the reader with further explanations or gives reasons for the observed differences in findings or unexpected outcomes. The occurrence of this move type is in line with Yang and Allison's (2003) study focusing on the Discussion section in applied linguistics RAs. In their study, this move/step is labelled as 'Accounting for results', providing readers with further explanations or giving the reasons for the observed differences in findings or unexpected outcomes. Examples of this move/step are shown below.

[10] (S13) *With regard to* ***the first factor***, ***it is probable that*** *the students have developed beliefs and expectations about education.*

MP 1

[11] (S3) *This result* ***may be explained by two different views***. (S6) ***One possible explanation is*** *that they were optimistic about their study, to view learning experience as success is kind of support and encourage willingness to learn better, even though* ***the present finding seems not to support (Ref)****'s study which pointed out that higher grades encouraged more willingness to work hard.*

HS 1

[12] (S11) *In regard to classroom practices,* ***this produced two avenues of explanation*** *for the researcher.* (S12) ***The first avenue would be that*** *any approach designed to develop students' communicative competence in order to communicate effectively in English serves as a corollary of CLT.*

JHNU 1

Some explanations of this particular communicative purpose, namely, explaining the results being discussed, are given through the use of words or phrases, i.e., *it is probable that*, *may be explained by, one possible explanation is that, this produced two avenues of explanation.* etc. Specifically, the present simple tense, be it active (*is probable that*) or passive construction (*may be explained*), and the modal *may* indicating possibility, can be used to accomplish this move/step. More importantly, as shown in [11], comparing results (Step 2) and explaining results (Step 3) can be embedded in one sentence, revealing that a possible explanation is reviewed with and supported by reference to the findings of previous studies.

Furthermore, it can be seen that the authors prefer to use modals of possibility such as *would* to express a hypothetical or conditional possibility as in [12]. As Nunan (2004) advocated, modality is the dimension of an utterance which allows the speaker or writer to reveal their attitudes towards 1) the propositional content, or 2) the illocutionary force of an utterance. Accordingly, authors may use modality to reveal their evaluative stance.

Move 2 Step 4: Presenting arguments/positions

One of the most common strategies to highlight results is by presenting positions. The authors use Move 2 Step 4 as a presentation of arguments to provide subjective judgment about the results and/or interpreting the findings being reported. This move/step includes information and interpretations that go beyond the objective results. The following exemplify the use of Move 2 Step 4 found in the dataset of this study.

[13] (S18) *Regarding the cultural appropriateness of CLT for Thai context, the researcher* ***reached the conclusion*** *that* ***there should be*** *an adapted version of CLT, taking into account Thai educational system*.

JHNU 1

[14] (S14) *The findings* ***suggest*** *one possible way that can help teachers to deal appropriately with low proficiency students*.

HS 1

[15] (S14) *Therefore,* ***it implies that*** *their closeness in age does not hinder the recognition of being a qualified and experienced instructor.*

HS 2

In the given examples, the researchers not only present the results, but also expound their ideas about them by making certain claims through expressions such as *reached the conclusion* and *there should be* to indicate either the certainty or tentativeness of their interpretation. In [14], the communication verb *suggest* in the present tense is also used for the purpose of stating extrapolations of the findings. Interestingly, [15] reveals that the authors can use the verb *imply* for the purpose of communicating an idea emerging from the findings of the study without saying it directly or to highlight generalisations based on the findings of the study.

Move 3: Evaluating the study

The communicative purpose of this move is to evaluate the study by stating both the strengths and weaknesses of the results. According to Yang and Allison (2003), this move was used least in the Discussion section of applied linguistics research articles, and no steps were used. In sharp contrast, analysis in the present study reveals that this move is used in all ten articles. This move consists of four steps, including stating the value of the study, offering pedagogic implications, indicating limitations, and suggesting further research.

Move 3 Step 1: Stating value of the study

The main function of Step 1 of Move 3 is to allow the authors to highlight the strengths or significance of the study which may be important to the research applications or implications of the findings. This move/step was present relatively rarely compared to other steps in this move, as it was found in only six (out of ten) research articles. The following instances illustrate how researchers stake their claims about the implications of findings.

[16] (S37) *To sum up,* ***this study provided a window into*** *the use and teaching practices of metacognitive and cognitive strategies by the Thai teachers of English at RMUTL*.

JHNU 2

[17] (S26) *Although the findings might not be transferable to other contexts with a high variation of students' characteristics since the data were collected from a small group of participants,* ***the present study is useful*** *in terms of the two aspects.*

HS 1

[18] (S7) *Despite some of limitations mentioned above,* ***the findings of the study provided various beneficial insights*** *for both foreign language teachers and students*.

HS 2

The instances above highlight the value, application, and advantages of the results of the studies. [16] shows the value of the use and teaching practices of metacognitive and cognitive strategies, whilst [17] claims that the results of the study are worth investigating with regard to learner attitudes. [18] indicates that the results are useful for foreign language teachers and students. With regard to the linguistic component, these instances reveal that, to announce the value and state the significance of the findings, statements in both present or past simple tense (*is, provided*) relating to the significance of the study, such as *provided a window into, various beneficial, useful, insights*, were by far the most common in this move/step. In all cases, deictic elements (*this, the*), followed by common nouns (*study, finding*), are also frequently employed to highlight certain useful discoveries from the study.

Move 3 Step 2: Offering pedagogic implications

Move 3 Step 2 allows the researchers to state the study implications. Since the small corpus analysed in this study was taken from the ten research articles in the field of applied linguistics, an applied discipline, most of the implications offered, therefore, are related to educational and pedagogical contexts. The examples below demonstrate this move/step.

[19] (S21) *This study* ***has significant pedagogical implications*** *for theory and practices*.

JHNU 1

[20] (S39) *The findings of this study* ***offered some practical implications for the teaching*** *of reading strategies.*

JHNU 2

[21] (S25) *Therefore, to improve their self-confidence in learning English, everyday English* ***might be introduced and trained*** *at the beginning through reading and listening activities, before productive language skills are incorporated.*

HS 1

As can be seen from the above, statements relating to the implications of the results to teaching and learning contexts (*pedagogical implications*) are commonly used to address Step 2 of Move 3. Linguistically, the modals of obligation/suggestion and possibility such as *might* are used in order to express author recommendation or to state explicit ways of using research outcomes in teaching practices. With respect to the presence of this move, the finding is consistent with Yang and Allison's (2003) study in that this move/step is reported and labelled as 'Drawing pedagogical implications', occurring in the Discussion section when this is the final section. Otherwise, it may be present in the Conclusion or Pedagogic Implications sections.

Move 3 Step 3: Indicating limitations

Move 3 Step 3 is followed by comments on the weaknesses of the results and serves to describe some limitations of the research presented. In this move/step, the researchers usually inform the readers that the reported study was limited, and thus subsequent interpretation and generalisation of the findings should be conducted with caution. Move 3 Step 3 was found in eight research articles of the entire dataset analysed. The realisation of this move/step is shown in the following excerpts.

[22] (S49*) Despite the interesting findings and their beneficial implications,* ***this study was limited*** *in several ways.* (S50) *Firstly, the reading strategies used in* ***this study were not exhaustive****.*

JHNU 2

[23] (S26) *Although* ***the findings might not be transferable to*** *other contexts with a high variation of students' characteristics since the data were collected from a small group of participants, the present study is useful in terms of the two aspects*.

HS 1

[24] (S5) *In the final analysis, the perspectives of these three FLTAs who participated in my study mirror particular stories of individuals in* ***this small setting****.* (S6) *Consequently,* ***different stories might be told by different groups of participants****, and they will also reflect various aspects of the U.S. classroom culture.*

HS 2

In examples [22] and [23], the limitations inherent in these studies are presented using the words and expressions (e.g., *limited, might not be transferable to*) in conjunction with reasons why the results should be interpreted with care (e.g., *the data were collected from a small group of participants*). Furthermore, the authors may implicitly inform readers about the limitations of the results by introducing a possible improvement to the methodology to provide more reliable results in future studies *(in this small setting, different stories might be told by different groups of participants*) as in [24]. Modal verbs expressing possibility (*might*) are also used in this move/step. The strategies used in this move/step allow the authors to caution readers that the study was not perfect, and thus generalisation of the findings should be conducted with care. However, it is interesting to note that this move/step does not appear in the studies of research articles in the social sciences conducted by Holmes (1997) or computer studies by Posteguillo (1999).

Move 3 Step 4: Suggesting further research

In Move 3 Step 4, the authors state some possible directions and areas for future research. This move/step, if found, closes the Discussion section. From the analysis, Move 3 Step 4 usually follows Step 3 of Move 3 (Indicating limitations). The examples below illustrate the features of this move/step.

[25] (S26) *As CLT is based in ESL context, its implication in EFL* ***is still subject to further research****.*

JHNU 1

[26] **(S30)** ***In future investigations, it might be possible to expand*** *the interesting findings from this current study on why low or high proficiency students perceive themselves as successful or unsuccessful.*

HS 1

[27] (S17) ***Future studies should also focus on*** *the process of culture learning in a particular language class in order to find out a practical approach for culture learning and teaching*.

HS 2

As shown in the above instances, the authors advocate the need for further research or offer recommendations about what should be done in the future. As can be seen from [25], [26], and [27], to signal and accomplish this purpose, this move/step makes frequent use of expressions such as *further research, future investigation*, and *future studies*. In this regard, the need for further research suggested is reinforced through the use of the modal of obligation/suggestion '*should*', to point out that in order to move the field forward, additional studies on the particular topic being discussed need to be conducted.

Discussion and Conclusion

This article presents an empirical study of the Discussion sections found in research articles in the discipline of applied linguistics. The findings of this study suggest that the rhetorical moves of this particular section include: '*Move 1: Reviewing the present study*,' '*Move 2: Analysing and extending from results*,' and '*Move 3: Evaluating the study*'. Each of the moves has very distinct rhetorical functions within this particular section. Each identified move type consists of a number of steps, with the exception of Move 1. Concerning their frequency of occurrence, at the move level, Move 1 was found in eight (out of ten) articles, whereas Moves 2 and 3 were found in every Discussion section of the ten articles analysed, suggesting that they play a crucial role in this particular section.

With respect to the rhetorical moves, the results of the analysis of the present study corroborate previous studies, including Basturkmen (2012) who analysed 10 discussion sections of dentistry research articles and Yang and Allison's (2003) study on applied linguistics research article discussions, in that the Discussion section usually begins with the review of the present study. This result is also in line with Kanoksilapatham (2012) who previously analysed the Discussion section of 60 engineering articles written in English, particularly in terms of the sequence of individual move types usually occurring in the articles. That is, three major move types found in their study consist of: *Move 1: Review the present study, Move 2: Consolidate results,* and *Move 3: State limitations and possible future research*.

It was found that the Discussion sections of the ten articles are likely to begin with Move 13: Reviewing the present study. This finding is in agreement with previous studies on this particular section (e.g., Basturkmen, 2012; Kanoksilapatham, 2012). That is, Kanoksilapatham's (2012) study focused on 60 discussions of three engineering sub-disciplines, whereas Basturkmen (2012) analysed 10 discussions of research articles published in the British Dental Journal. In this regard, although these two studies focused on the Discussion sections of different disciplines, they seem to share the common generic identity in the Discussion section with regard to facilitating readers access to background information of the research before discussing the findings of the study being reported, without going back to information already established in the Introduction and/or the Methods sections.

However, the model for the Discussion section proposed in the current study differs from the one proposed by Yang and Allison (2003), in terms of both the move types and step levels, even though their model was developed from the analysis of research articles in the discipline of applied linguistics. Moreover, this finding is partly in agreement with Swales and Feak (2004) when describing typical moves in the Discussion section of academic research papers. They suggest that the Discussion section consists of five moves: '*Move 1: Restate the findings and accomplishments*,' '*Move 2: Evaluate how the results fit in with the previous findings*,' '*Move 3: List potential limitations to the study*,' '*Move 4: Offer an interpretation/explanation*,' and '*Move 5: State the implications and recommend further research*'. At this juncture, due to the small corpus of the present study, this finding remains inconclusive. Likewise, as these studies analysed research articles from different academic disciplines, the relevance of disciplinary variations, therefore, is still highly possible and needs to be taken into account.

In view of the pedagogical significance of studying linguistic features in relation to communicative functions, as the findings of this study show, it is acknowledged that communication verbs such as *suggest*, *provide*, and modality such as *could, should, might* are generally used in various parts of the Discussion section. This is because RA authors need to generalise and interpret the findings of the study being presented. Therefore, the functions of these linguistic features need to be highlighted, especially while teaching these lexical and syntactical features to students who need to get their research papers published.

To enhance the opportunities for scholarly publication, understanding the rhetorical organisation of research articles is crucial to conform to the expectations of the target discourse community. As far as pedagogical implications are concerned, students need to be made aware of the schematic patterning of research articles and the conventions established by the respective discourse community. They should be instructed to see the rhetorical complexities and relationships among communicative functions identified in each move and step and the use of linguistic features. The results generated by this study can help inform strategies in teaching the reading and writing of research articles in order to maximise the opportunities for scholarly publication for novice scholars and graduate students in Thailand.

Instructors can teach students, particularly those at an advanced level, that when they read or write research articles, they can expect to encounter rhetorical patterns and linguistic features and variations prevalent in each section of research articles. Pedagogically, for example, in a class, instructors may carefully select and assign students to read samples of academic writing, including introductions, methods, and discussions. At the initial stage, in pairs or small groups, the students could be asked to analyse the text and try to list the stages found in these sections. Instructors may encourage the students to consider the texts by asking them to answer the following questions: "What is the author trying to tell us?", "What kind of information is included in the texts?", "What kind of information is likely to be found at the beginning and end of the text?", "What is the tone of language that the author uses, informative or persuasive?", and "What are some of the linguistic features found in the texts?". During this period, lecturers can move about the room providing guidance and answering queries.

After the students are familiar with the written texts, lecturers may make use of the findings of this study, for instance, to demonstrate and highlight the existence of the schematic structure, linguistic features, and their communicative functions realised in a research article's discussion section. Then, to encourage the students to become more involved in the class, it might prove intriguing and challenging to assign the move/step identification task, by directing students' attention to focus on a set of linguistic features prevailing in the text. Making students aware of the rhetorical structure and linguistic characteristics that they are likely to encounter when reading or writing research articles can help prepare them to better cope with a tremendous range of language variation, especially when they are encouraged to publish the results of their research in referred journals, particularly those indexed in the TCI database.

References

Basturkmen, H. (2012). A genre-based investigation of discussion sections of research articles in dentistry and disciplinary variation. *English for Academic Purposes, 11*(2), 134-144.

Bhatia, V.K. (1993). *Analysing genre: Language use in professional settings*. London: Longman.

Biber, D., Conner, U., Upton, T., & Kanoksilapatham, B. (2007). Introduction to move analysis. In D. Biber, U. Conner, & T. Upton (Eds.), *Discourse on the move: Using corpus analysis to describe discourse structure* (pp. 25-41). Amsterdam: John Benjamins Publishing Company.

Brett, P. (1994). A genre analysis of the results section of sociology articles. *English for Specific Purposes, 13*(1), 47-59.

Chalapati, S. (2007). *The internationalisation of higher education in Thailand: Case studies of two English-medium business graduate programs* (Unpublished PhD dissertation). RMIT University, Australia.

Coleman, J.A. (2006). English-medium teaching in European higher education. *Language Teaching, 31*(1), 1-14.

Conner, U., & Upton, T. (2004). The genre of grant proposals: A corpus linguistic analysis. In U. Conner, & T. Upton (Eds.), *Discourse in the professions: Perspectives from corpus linguistics* (pp. 235-256). Amsterdam: John Benjamins.

Cook, G. (1990). *Discourse*. Oxford: Oxford University Press.

Crystal, D. (2004). The past, present, and future of World English. In A. Gardt, & B. R. Huppauf (Eds.), *Globalization and the future of German* (pp. 27-45). Berlin: Mouton de Gruyter.

Ding, H. (2007). Genre analysis of personal statements: Analysis of moves in application essays to medical and dental schools. *English for Specific Purposes, 26*, 368-392.

Dudley-Evans, A. (1994). Genre analysis: an approach for text analysis for ESP. In M. Coulthard (Ed.), *Advances in written text analysis* (pp. 219-228). London: Routledge.

Flowerdew, J. (2001). Attitudes of journal editors to non-native speaker contributions. *TESOL Quarterly, 35*(1), 121-150.

Holmes, R. (1997). Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in the three disciplines. *English for Specific Purposes, 16*(4), 321-337.

Hyland, K. (2004). *Genre and second language writing*. Ann Arbor, MI: University of Michigan Press.

Kanoksilapatham, B. (2007a). Writing scientific research articles in Thai and English: Similarities and differences. *Silpakorn University International Journal, 7*, 172-203.

Kanoksilapatham, B. (2007b). Rhetorical moves in biochemistry research articles. In D. Biber, U. Conner, & T. Upton (Eds.), *Discourse on the move* (pp. 73-120). Amsterdam: John Benjamins.

Kanoksilapatham, B. (2011). Language of civil engineering Introductions: Textual structure and linguistic

characterizations. *The Asian ESP Journal, 7*(2), 55-84.
Kanoksilapatham, B. (2012). In search of the generic identity of the discussion section: Three engineering sub-disciplines. *Taiwan International ESP Journal, 4*(2), 1-26.
Kanoksilapatham, B. (2015). Distinguishing textual features characterizing structural variation in research articles across three engineering subdiscipline corpora. *English for Specific Purposes, 37*(1), 74-86.
Laurence, A. (1999). Writing research article introductions in software engineering: How accurate is a standard model? *IEEE Transactions on Professional Communication, 42*(1), 38-46.
Lim, J.M.H. (2006). Method sections of management research articles: A pedagogically motivated qualitative study. *English for Specific Purposes, 25*(3), 282-309.
McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge: Cambridge University Press.
Nunan, D. (2004). *Introducing to discourse analysis*. Quarry Bay: The Central Printing Press.
Nwogu, K. (1997). The medical research paper: Structure and functions. *English for Specific Purposes, 16(*2), 119-138.
Ozturk, I. (2007). The textual organization of research article Introductions in applied linguistics: Variability within a single discipline. *English for Specific Purposes, 26*(1), 25-38.
Paltridge, B. (1994). Genre analysis and the identification of textual boundaries. *Applied Linguistics, 15*(3), 288-299.
Pang, T. (2002). Textual analysis and contextual awareness building: A comparison of two approaches to teaching genre. In A. Johns (Ed.), *Genre in the classroom: Multiple perspectives* (pp. 145-162). New Jersey: Lawrence Erlbaum.
Peacock, M. (2011). The structure of the Methods section in research articles across eight disciplines. *The Asian ESP Journal, 7*(2), 99-122.
Posteguillo, S. (1999). The schematic structure of computer science research articles. *English for Specific Purposes, 18*(2), 139-160.
Samraj, B. (2002). Introductions in research articles: Variation across disciplines. *English for Specific Purposes, 21*, 1-17.
Sinlarat, P. (2004). Thai universities: Past, present, and future. In P.G. Altbach, & T. Umkoshi (Eds.), *Asian universities: Historical perspectives and contemporary challenges* (pp. 201-220). Baltimore: The Johns Hopkins University Press.
Svasti, J., & Asavisanu, R. (2007). Aspects of quality in academic journals: A consideration of the Journals published in Thailand. *Science Asia*, *33*, 137-143.
Swales, J.M. (1990). *Genre analysis: English in academic and research setting*. Cambridge: Cambridge University Press.
Swales, J.M. (2004). *Research genre: explorations and applications.* Cambridge: Cambridge University Press.
Swales, J.M., & Feak, C.B. (2004). *Academic writing for graduate students: Essential tasks and skills*. Ann Arbor: University of Michigan Press.
Van Mulken, M., & van der Meer, W. (2005). Are you being served? A genre analysis of American and Dutch company replied to customer inquiries. *English for Specific Purposes, 24*(1), 93-109.
Yang, R., & Allison, D. (2003). Research articles in applied linguistics: Moving from results to conclusion. *English for Specific Purposes, 22*(4), 365-385.

Establishing a Long-Term Reading Habit in English as a Foreign Language: A Case Study

Kyung Sook Cho*
Busan National University of Education, South Korea

Stephen Krashen
University of Southern California (Emeritus), USA

Abstract

This paper reports a case study to examine the effects of long-term reading habits on second language acquisition. The case reported helps inform us about how learners of English as a foreign language can continue to improve in English on their own. The learner, Jung Seo, while living in Korea, developed a long-term reading for pleasure habit in English, utilizing strategies such as narrow reading, finding compelling reading material, finding a source of books, and finding a time and place to read. The participant reported remarkable progress in their English language development.

Keywords: EFL, Pleasure Reading, A Long- Term Reading Habit, Narrow Reading, Language Development

Introduction

According to Krashen (2004a), reading is the primary source of much of our literacy development and our knowledge of the world; however, it takes a great deal of reading over many years to experience these benefits.

We address here what we consider to be the crucial question in research on self-selected reading: How can we help those trying to develop their language skills to establish a long-term reading habit? Our focus here is adult second language acquisition in the foreign language context.

We present the case of Jung Seo, now 36 years old, who became interested in reading as the result of a short introduction to the research, established a reading habit that so far has lasted for more than four years, and has reported satisfying progress in English language development while living in Korea. Jeong Seo's experience confirms several important hypotheses related to the impact of free voluntary reading on language development, the factors required for the establishment of a long-term reading habit, and the use of effective reading strategies.

How It Began

Jung Seo was an English major as an undergraduate, but the program was traditional, and did not include any self-selected or light reading. After graduation, she worked as an elementary school teacher for several years. Jung Seo was required to teach English, but felt that she was not well enough prepared. She described her English as "fumbling" and said she made a lot of mistakes when speaking and paused a lot to think before saying anything.

She then went to graduate school, and was exposed to the research on self-selected reading in one of the courses they were taking. This inspired her to begin reading on her own in order to improve her English.

**Corresponding author. Tel. 820515007315. Email: kscho@bnue.ac.kr, Busan National University of Education, Department of Elementary English, 34 Gyo-de Ro, Yeon-je Gu, Busan, South Korea*

The Learner's Progress

Ms. Seo had never before experienced reading for pleasure in English, but, at the time of this writing, has been reading for the last four years and eight months. She typically reads one hour a day (but sometimes more on weekends, see below) and has read about 200 books in English (190 adult novels, 10 young adult novels) since starting her program, and about 600 children's books. Jung Seo has reported that her improvement in English has been "dramatic."

Language Development

Reading Comprehension

Ms. Seo can now read easily in English, and has no trouble with books that she had great difficulty with when she tried to read them several years ago. In fact, she now prefers to read books in the original English rather than in Korean translation, and she says that she reads English novels as easily as she reads Korean novels:

> I recently read a book I had read one year ago with difficulty and hardly thought about the fact that it was in English. I read it as fluently as I read in Korean. (translated from interview in Korean with Ms. Seo by KS Cho).

Less Use of the Dictionary

Ms. Seo is far less dependent on the dictionary when reading than she was when first beginning to read in English. In common with many adult second language learners, Ms. Seo found it difficult to free herself from constant use of the dictionary:

> I had a hard time getting rid of my old habit of frequent trips to the dictionary, but ...all of a sudden I was able to focus on only the content of the words. (translated from interview in Korean with Ms. Seo by KS Cho)

This suggests greater vocabulary knowledge thanks to reading, but there is another explanation, discussed below.

Oral Language

Ms. Seo feels she is now much more fluent in spoken English than she was a few years ago: her "fumbling time decreased a lot when talking" (translated from interview in Korean with Ms. Seo by KS Cho) and she no longer hesitates to speak to native speaker teachers of English during meetings.

Listening Comprehension

Ms. Seo reports that she can now understand American soap operas without captions if she has some background knowledge about the story and can understand English language movies without subtitles:

> I can now understand American soap operas without captions if I have some background knowledge about the story and can understand English language movies without subtitles. (translated from interview in Korean with Ms. Seo by KS Cho)

Improvement in Attitude toward Reading

Ms. Seo clearly enjoys reading in English, which was not true in the past:

> ... above all, the most valuable effect of pleasure reading to me is, as the word "pleasure reading" indicates, is that English reading now is really a pleasure. (translated from interview in Korean with Ms. Seo by KS Cho)

Learner's Progress after Four Years and eight Months of Reading

Jung Seo's progress is very close to that predicted by Nation (2014) and Krashen and Mason (2015): On the basis of several case histories collected by Mason and Krashen (2015) concluded that an adult additional language learner of English could move from the low intermediate level (TOEIC 250) to a very advanced level (TOEIC 950) in about 1220 hours of self-selected pleasure reading. This is nearly exactly what Jung Seo did.

Reading Strategies

In addition to validating "the power of reading" (Krashen, 2004a) and the Comprehension Hypothesis (Krashen, 2014), Seo's report provides evidence of the value of several strategies used by successful language learners.

Narrow Reading

It has been argued that narrow reading is best for language and literacy development. This means reading several books by one author or about a single topic of interest (Krashen, 2004b). Narrow reading and narrow input in general provide natural repetition of grammar, vocabulary, and aspects of style, as well as facilitating comprehension because the context is familiar to the reader and becomes even more familiar as the reader reads more. There is some evidence that good readers are narrow readers (Lamme, 1976) and previous case histories have provided additional evidence that it has a positive effect (Cho & Krashen, 1994, 1995).

Ms. Seo became a narrow reader. At first, she chose best-sellers and books that had been made into movies. She soon discovered her favorite authors: First was Sidney Sheldon, and then it was Sophie Kinsella. Seo reports that when she discovered Kinsella's books, she "could not stop reading them" (translated from interview in Korean with Ms. Seo).

She read her first Kinsella novel in one day, starting in the morning on the subway, then more when she had time at school, and then after she came home until 3 am. She is now reading her seventh Michael Connelly Harry Bosch novel, having read the previous six in the series.

Finding Compelling Comprehensible Input

It has been hypothesized that the most effective input for language development is more than being just interesting; it is compelling, and puts the language acquirer in a state of "flow" (Csikszentmihalyi, 1992; Lao & Krashen, 2008). Compelling reading is so interesting that the acquirer "'forgets' that they are reading in another language and become "lost in the book" (Nell, 1990). When input is compelling, there is less concern about "noise" in the input and about pockets of incomprehensible input (e.g., unfamiliar vocabulary), because the reader is so intent on understanding the overall message.

Ms. Seo sought out compelling reading material, and clearly found it, as evidenced by her absorption in Sophie Kinsella's novel (see above) and her continuous reading over more than four years. She has also made it clear that her reading is a source of great joy: "Reading has given me a happier life..." (translated from interview in Korean with Ms. Seo by KS Cho) and she reports that she greatly enjoys her commuting time because it is also reading time (see below).

Her observation that as time went on she had less inclination to use the dictionary could be a result of the input being so compelling that she ignored the distractions, or her growing vocabulary competence, thanks to reading, or both. Evidence for the former hypothesis is her statement that: " If a book was interesting enough, I was able to read despite difficult words" (translated from interview in Korean with Ms. Seo by KS Cho).

Finding a Source of Books

Ms. Seo had an important advantage that not all second language acquirers have: A convenient source of books. Ms. Seo reported that her major source of books was the local English library, which had "a variety of different books" (translated from interview in Korean with Ms. Seo by KS Cho). When she began her reading journey, she had a hard time finishing a book before the due date, but eventually she was checking out two to three books every two weeks.

Finding a Time and Place to Read

Readers need time and a space to read, often difficult to find in these hectic times. Ms. Seo solved the problem by reading in the morning on the subway during her 20-minute ride to work and another 20 minutes on the return home. In the beginning, she reported that she could only read a few pages before it was time to get off the subway, but this changed as her competence grew.

In addition to the subway, she reads before going to bed, during recess at the school where she teaches, on trips, and, as mentioned earlier, at home on weekends and during vacations (sometimes reading all day long).

Conclusion and Discussion

Reading has not been the only source of English input for Jung Seo during the last four years. She has had occasional conversations with native speakers and has seen movies and TV shows, but these sources provided nowhere near the amount of input she received from reading. She gives reading the credit for her progress. Also, during the last four years, she did not engage in English "study" either formally or informally.

We did not administer standardized tests to Jung Seo, but her story is noteworthy. Also, we did not compare her progress to subjects who spent an equal amount of time in voluntary study. We suspect such individuals are rare. In contrast to the case histories of successful first and second language development through pleasure reading (e.g., Krashen, 2004a), there are cases of intensive and prolonged study that resulted only in frustration (see Krashen, 2014, for a description of Gouin's experiences).

It can, of course, be argued that there is an optimal mix of comprehensible input and study, Jung Seo's experiences confirm, however, that substantial progress in a second language can take place through reading, and without pain. In fact, it can be extremely pleasant, so pleasant that some people will gladly continue doing it, which helps to ensure continuous progress.

References

Cho, K., & Krashen, S. (1994). Acquisition of Vocabulary from the Sweet Valley Kids series: Adult ESL Acquisition. *Journal of Reading*, 37, 662-667.

Cho, K., & Krashen, S. (1995). From Sweet Valley Kids to Harlequins in one year. *California English*, 1(1), 18-19.

Csikszentmihalyi, M. (1992). *Flow: The psychology of optimal experience*. New York: Harper Perennial.

Krashen, S. & Mason, B. (2015) Can second language acquirers reach high levels of proficiency through self-selected reading? An attempt to confirm Nation's (2014) results. *International Journal of Foreign Language Teaching,* 10(2), 10-19.

Krashen, S. (2004a). *The Power of Reading.* Portsmouth: Heinemann, Westport, Conn: Libraries Unlimited.

Krashen, S. (2004b). The case for narrow reading. *Language Magazine,* 3(5),17-19.

Krashen, S. (2014). Case Histories and the Comprehension Hypothesis. *TESOL Journal* (www.tesol-journal.com), June, 2014.

Lamme, L. (1976). Are reading habits and abilities related? *Reading Teacher* 30, 21-27.

Lao, C. & Krashen, S. (2008). Heritage language development: Exhortation or good stories? *International Journal of Foreign Language Teaching,* 4(2), 17-18.

Nation, I.S.P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language*, 26(2), 1-16.

Nell, V. (1990). *Lost in a book*. New Haven, CT: Yale University Press.