



# TESOL International Journal

---

*Teaching English to Speakers of Other Languages*

Volume 15 Issue 2 2020

ISSN 2094-3938

Published by the TESOL International

Journal <http://www.tesol-international-journal.com>

© English Language Education  
Publishing Brisbane Australia

This book is in copyright. Subject to statutory exception no reproduction of any part may take place without the written permission of English Language Education Publishing.

**No unauthorized photocopying**

All rights reserved. No part of this book may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise, without the prior written permission of English Language Education Publishing.

Chief Editors: Dr. Custódio Martins

Ramón Medriano, Jr.

**ISSN. 2094-3938**

# TESOL International Journal

## Chief Editors

Custódio Martins  
*University of Saint Joseph, Macao*

Ramón Medriano, Jr.  
*Pangasinan State University – School of Advanced Studies*

## Senior Associate Editors

Jun Zhao  
*Augsuta University,  
USA*

Peter Ilić  
*The University of Aizu,  
Japan*

Farzaneh Khodabandeh  
*Payame Noor University,  
Iran*

## Associate Editors

Mário Pinharanda Nunes  
*University of Macao, China*

Sharif Alghazo  
*University of Jordan,  
Jordan*

Khadijeh Jafari  
*Gorgan Islamic Azad  
University, Iran*

Rining Wei  
*Xi'an Jiaotong-Liverpool University, China*

Harriet Lowe  
*University of Greenwich,  
UK*

## Editorial Board

**Abdel Hamid Mohamed** - Lecturer, Qatar University, Qatar

**Adriano, Nina** - Baliuag University, Philippines

**Al-Dhaif, Amina** - Northumbria University, UK

**Alhilali, Tayba** - Lecturer, Higher College of Technology, Sultanate of Oman

**Badwan, Khawla** - Manchester Metropolitan University, UK

**Baker, John** - Ton Duc Thang University, Ho Chi Minh, Vietnam

**Balchin, Kevin** - Canterbury Christ Church University, UK

**Bekteshi, Edita** - University of Tirana, Albania

**Boonsuk, Yusop** - Prince of Songkhla University, Thailand

**Çakir, İsmail** - Yıldırım Beyazıt University, Turkey

**Chan, Chun Chuen** - University of Sydney, Australia

**Chen, Qi** - Newcastle University, UK

**Chung, Hiu Yui** - The Open University of Hong Kong, China

**Cutrone, Pino** - Nagasaki University, Japan

**Derakhshan, Ali** - Golestan University, Gorgan, Iran

**Dodigovič, Marina** - Universidad de La Rioja, Spain

**Essex, James** - Waseda University, Japan

**Farsani, Mohammad Amini** - Iran University of Science and Technology, Iran

**Geden, Ayşe Gür** - University College London, UK

**Ghannam, Jumana** - Nottingham Trent University, UK

**Hajan, Bonjovi H.** - José Rizal University, Philippines

**Hasan, Idrees Ali** - American University of Kurdistan, Duhok, Kurdistan Region

**Hos, Rabia** - Rhode Island University, USA

**Kambara, Hitomi** - The University of Texas Rio Grande Valley, USA

**Kazemian, Mohammad** - Guilan University of Medical Sciences, Iran

**Ku-Mesu, Katalin Egri** - University of Leicester

**Lin, Yu-Cheng** - The University of Texas Rio Grande Valley, USA

**Maher, Kate** - Kyoto University of Foreign Studies, Japan

**Mohamed, Naashia** - University of Auckland, New Zealand

**Munalim, Leonardo O.** - Philippine Women's University, Philippines

**Mustafa, Faisal** - Syiah Kuala University, Banda Aceh, Indonesia

**Niu, Ruiying** - Guangdong University of Foreign Studies, China

**Rozells, Diane** - Sookmyung Women's University, South Korea

**Salem, Ashraf** - Sadat Academy for Management Sciences, Egypt

**Saito, Akihiro** - Hakuoh University, Oyama, Japan

**Sakka, Samah Mohammed Fahim El** - Suez University, Egypt

**Slaght, John** - University of Reading, UK

**Stewart, Alison** - Gakushuin University, Japan

**Tzu-Shan Chang** - Tamkang University, Taiwan

**Ulla, Mark** - Walailak University, Thailand

**Venela, R.** - National Institute of Technology, Warangal, India

**Wong, Kevin** - Pepperdine University, Los Angeles, USA

**Yuanhua Xie** - Guangdong University of Foreign Studies, China

**Yusri, Y.** - Fajar University, Makassar, Indonesia

**Zayani, Emna Maazoun** - Sfax and Exeter University, UK

## Foreword

The articles in the current issue of the TESOL International Journal explore and discuss the pedagogical implications of writing, speaking and types of testing for the language classroom.

**Bayley and Lee's** study focuses on the use of Grammarly as a new tool to foster second language writing, providing learners with the necessary support for writing tasks, and promoting "motivation and confidence". The authors demonstrate that Automatic Writing Evaluation tools, namely Grammarly, can detect errors in writing almost as effectively as teachers. Through the analysis of three different corpora and four genres, the study compares types of errors, their frequency, lexical and syntactic variation. One of the findings of Bayley and Lee's study was that "Grammarly was successful at identifying local level errors in L2 writing regardless of the writing genre." Notwithstanding this finding, Bayley and Lee argue that even if Grammarly is adequate to identify what they define as "surface-level errors", such as articles or prepositions, the role of the language teacher is vital "(...) for issues related to awkward wording and cohesion (...)".

The next article by **Jun Zhao**, on academic writing, reports a pedagogical trial in which students learned reporting verbs in a composition course. In the course, students have to produce argumentative research pieces of work in which they have to be able to integrate other sources. The main goal of this trial was to understand whether explicit teaching of reporting verbs had any effect on the final writing product. The results of this pedagogical trial revealed that explicit teaching of reporting verbs had positive effects on the students' essays.

**Yi-Ching's** study examines the extent to which classroom assessment impacts learning from the students' point of view of summative and formative assessment. Yi-Ching administered a questionnaire to university students in Taiwan based on "Dorman and Knightley's (2006) PATI and Green's (2007) model of washback". The analyses of the data revealed that even if students prefer summative assessment, they also consider that the combination of both summative and formative assessment brings benefits to their learning process.

**Marzieh's** study addresses L2 pragmatic development from a sociocultural perspective. The main goal of the study was to compare "individual and collaborative languaging" based on three L2 output functions: "noticing, metalinguistic reflection and hypothesis testing". The results of the study show differences between individual and collaborative oral language production. Even though both individual and collaborative languaging enhance noticing and metalinguistic reflection, collaborative production has a more significant effect in terms of hypothesis testing.

Custódio Martins

University of Saint Joseph, Macao

## Contents

Foreword	1
An Exploratory Study of Grammarly in the Language Learning Context: An Analysis of Test-Based, Textbook-Based and Facebook Corpora <i>Daniel Bailey</i> <i>Andrea Rakushin Lee</i>	4
Teaching Reporting Verbs to English as a Second Language Undergraduate Writers in the Academic Context <i>Jun Zhao</i>	28
Taiwan University Students' Perceptions of Summative and Formative Classroom Assessment in English Courses <i>Yi-Ching Pan</i>	46
A Comparative Study of Individual and Collaborative Oral Languageing for L2 Speech Act Production <i>Marzieh Bagherkazemi</i>	65

## An Exploratory Study of Grammarly in the Language Learning Context: An Analysis of Test-Based, Textbook-Based and Facebook Corpora

**Daniel Bailey**

*Konkuk University Glocal Campus, South Korea*

**Andrea Rakushin Lee\***

*Konkuk University Glocal Campus, South Korea*

### Abstract

Different genres of writing entail various levels of syntactic and lexical complexity, and how this complexity influences the results of Automatic Writing Evaluation (AWE) programs like Grammarly in second language (L2) writing is unknown. This study explored the use of Grammarly in the L2 writing context by comparing error frequency, error types and writing complexity for university admission test essays, textbook-based descriptive essays, social network site (SNS) posts, and SNS comments. Several findings were revealed. Punctuation, grammar, vocabulary, and spelling mistakes were the most common error types across all corpora. In particular, determiner errors such as article usage were most frequent while differences existed across corpora related to incomplete sentences, run-on sentences, and spelling. Test-based and textbook-based compositions consisted of longer sentences with less lexical variation than SNS-based writing. SNS posts and comments produced greater noun, verb, and modifier variation. After applying Grammarly corrections, SNS posts resulted in the greatest clarity and this was attributed to shorter sentence length and simpler word-choice than textbook-based and test-based writing. Grammarly was more appropriate for local surface-level errors (e.g. articles, preposition, and verb-noun agreement) while instructors are needed for issues related to awkward wording and cohesion. Pedagogical implications in the English as a Foreign Language (EFL) context are given.

**Keywords:** corrective feedback, automatic writing evaluation, Grammarly, second language writing, syntactic complexity, and lexical density, Facebook

---

\* Email: andrealee1216@gmail.com.

## Introduction

While controversy currently exists concerning the application of Automatic Writing Evaluation (AWE) in test settings, there is little argument to the time-saving benefits afforded by computer generated feedback. The accuracy of AWE programs increases each year, and there is a growing debate among language instructors on whether or not to allow students to use these tools to improve their L2 writing. Research has found that AWE programs are adequate at detecting errors (Burston, 2008) at rates approaching the effectiveness of teacher corrections (Nadasdi & Sinclair, 2007). AWE programs can increase linguistic writing quality and engagement (Gauthier, 2013). This observed improvement in accuracy is accompanied by heightened levels of motivation and confidence (Potter & Fuller, 2008). Students are satisfied with corrective feedback from programs like Grammarly, especially in conjunction with a human rater (O'Neill & Russel, 2019), and while research shows the utility of programs like Grammarly, exactly how AWE programs perform across different L2 writing genres or error types is still unclear.

Automatic writing evaluation has undergone scrutiny over the years. Critics cite shortcomings with AWE concerning the poor validity of scoring tests (Chung & Baker, 2003) and unreliable assessment of social communication (Ericsson, 2006). Further condemnation persists because AWE programs have difficulty recognizing deeper level, global errors, related to such things as cohesive links and errors in factual content (McGee, 2006). Moreover, overreliance on AWE software prevents students from engaging language learning strategies such as looking up unknown words or asking for help. According to the comprehensible output hypotheses, what learners do when pushed is responsible for some amount of language acquisition (Swain, 2000). Language learning occurs in part by clarifying meaning during the writing process, but this clarification entails less cognitive engagement with AWE. Consequently, students spend less time looking up unknown words or discussing corrective feedback with more knowledgeable peers.

Despite the opposition, there are undeniable advantages for L2 writers and instructors when implementing AWE. The accuracy of AWE programs continues to improve in parallel with decreasing costs, and time saved by teachers can be allocated to other student needs. Furthermore, student-centered feedback with AWE addresses some of the challenges common in differentiated EFL/ESL instruction. Scaffolding feedback, according to language proficiency, allows the L2 writer to negotiate meaning within their zone of proximal development. Krashen's input hypothesis (1985) posits that language learners progress in their knowledge of the language when they comprehend language input slightly above their current level. AWE feedback provides insight into the learner's next stage of language acquisition because the revisions address what the user composed on their own. Students receive input from AWE at an  $i+1$  level, where  $i$  is learner's interlanguage and  $+1$  is the next stage of language acquisition. The benefits of AWE depend on how writing is defined and the content that is written (Ware, 2011).

Understanding the efficacy of AWE platforms with different writing conditions helps justify its use in the EFL context. Factors such as L2 writing anxiety, available time to write, academic stakes,



and the audience are considered here to be conceptually unique for test-based, textbook-based, and SNS-based writing genres. Therefore, text from each of these genres is analyzed.

The test-based corpus in the current study came from university entrance exam essay questions. Academic stakes are higher with test-based writing, and there is less time for pre-writing, when-writing, and post-writing strategies. Time limits with test-based writing prevent lengthy pre-writing planning strategies and post writing reflective strategies. High stakes conditions with writing tests also influence L2 writing anxiety.

Textbook-based writing produced in the classroom using time-consuming writing strategies results in linguistic differences when compared to other forms of composition. Textbook-based writing provides more time for more L2 writing strategies (e.g., mind-mapping, brainstorming, outlining, and reflection) and allows students to use models of correct composition. However, lower stakes with textbook-based writing may mean less motivation to perform well on writing tasks compared to university admissions test conditions.

Social network sites like Facebook allow students to practice real-world writing when communicating. The terms Facebook and SNS are used interchangeably throughout this manuscript and refer to informal real word written communication for social purposes. Students share their composition with SNS writing, and this elicits motivation by comparison but also anxiety from fear of looking foolish in front of their peers. SNS writing on platforms like Facebook can be categorized as main posts and comments/replies (*hereafter* comments). Main posts are similar to diary or journal entries while comments often entail shorter statements. Bailey et al. (2017) documented the patterns in participation among students in a Facebook for language learning program and recognized differences in sentence complexity between main posts and replies. The different parameters with each of these writing tasks are expected to produce different error types, error rates, and writing complexity.

Grammarly, an AWE platform that identifies over 300 types of writing errors, was used in the current study to compare error types and writing complexity for different EFL writing genres. No argument is being made that Grammarly is better than other AWE platforms. This study provides insight into the usefulness of AWE programs like Grammarly as an L2 writing tool by recognizing which types of errors Grammarly correctly identifies, and how those identified errors vary across different EFL writing tasks.

### Literature Review

AWE can assist both low- and high-performing students with their accuracy and save instructors valuable time. The current study posits that AWE programs like Grammarly are valuable L2 writing tools that can help educators grade writing and help students revise compositions. The following literature review explores the current state of AWE research and the role AWE has in the context of L2 writing.

### Automatic Writing Evaluation

The development of AWE software has enabled L2 learners to receive feedback on language and content in addition to automated scores. Benefits to writing accuracy were recognized in AWE studies (Anson, 2006; Dikli, 2010). These results drive the popularity of integrating automatic corrective feedback in the classroom (Li, Link, & Hegelheimer, 2015; Zhang & Hyland, 2018). Researchers continue to find AWE helpful for measuring linguistic accuracy (Li et al., 2015). Li et al. (2015) used AWE error reports to explore the role of Criterion in an ESL writing curriculum. By looking at error types, they found that automatic writing evaluation software led to increased revisions and improved accuracy. Li et al. (2015) required students to meet a benchmark score through AWE, which resulted in higher participation with writing practice and a heightened motivation to write.

AWE has the potential to liberate teachers by freeing up time; however, the capacity of AWE to provide useful feedback is still under debate (Anson, 2006). Stevenson and Phakiti (2014) set out in a meta-analysis of AWE research carried out in writing classrooms. Results indicated only modest evidence that AWE feedback had a positive effect on the quality of the texts students produced and little evidence about whether AWE was associated with more general improvements in writing proficiency. They suggest AWE feedback is more beneficial when combined with teacher feedback, and make a call for future research to examine AWE effectiveness in ESL and EFL settings. Li et al. (2015) make a similar call for research by noting a lack of evidence for the use of AWE corrective feedback in ESL writing classrooms.

Most AWE research has focused on a written product with less attention to the revision process (Zhang & Hyland, 2018). To address this gap in research, El Ebyary and Windeatt (2010) investigated whether the use of computer-based feedback using the AWE program Criterion would influence attitudes towards feedback, the writing process, and writing product. Through pre/post-survey analysis, interviews, and holistic scores from Criterion, they found that students used few pre-planning strategies and perceptions to feedback improved after using Criterion. The researchers recognized that teacher intervention is critical for motivation and program success and recommend the integration of AWE into conventional writing instruction. Drawbacks to AWE in their program included cheating and stagnant levels of revisions.

Zhang and Hyland (2018) conducted a study of two Chinese students by comparing AWE (with Criterion) and teacher feedback over the course of a semester. They found that AWE provided by Criterion tended to highlight errors instead of correcting them. Explicit corrective feedback has shown to be superior to simply bringing awareness to errors through indirect feedback (Chandler, 2013; Katayama, 2007), but simple awareness of the location of mistakes still leads to better writing outcomes (Ferris, 2004). AWE with Criterion provided a substantial amount of marginal comments, which could be considered a praising attribute because identifying errors is an early step to overcoming similar mistakes in future writing. However, the two Chinese students in Zhang and Hyland's (2018) study felt the automatic feedback was overwhelming and repetitive.

A current debate exists around AWE regarding effectiveness for students at different L2

proficiency levels. A certain amount of minimum proficiency is necessary to negotiate feedback delivered in the L2, but where that level rests is unknown. AWE systems may benefit more competent L2 writers more than less competent ones because stronger writers have the communicative and rhetorical understanding of the target language allowing them to make more appropriate use of the AWE feedback. In line with this argument, Dikli (2010) suggests some AWE is too overwhelming for low-performing writers because such writers may require basic metacognitive knowledge (Liao, 2016) or possess a base knowledge of the L2 (Caveleri & Dianati, 2016; Hoang & Kunnan, 2016). Other researchers have found AWE benefited low-performing writers (Chen & Cheng, 2008; O'Neill & Russell, 2019) by giving them input on correct writing form.

Some criticisms of AWE relates to a lack of human interaction. Dikli (2010) recommends adding mechanisms to AWE that promote individualized feedback, such as providing formative and cumulative feedback and increasing the number of praise messages. Dikli (2010) explored the nature of feedback that ESL students received on their writing either from a human rater or AWE. Teacher feedback was shorter and more to the point while AWE feedback was redundant (e.g., repetitive information) and not student centered. Feedback from the teacher was late or non-existent. According to Dikli (2010), it is necessary to consider effective methods to analyze large number of essays and provide individual feedback. The current study posits that AWE should be used to address local error types, such as surface-level revisions (Stevenson, 2016), while teacher feedback can focus on global issues.

Stevenson (2016) investigated the ways AWE is used or could be used as an instructional tool in the writing classroom and provides an overview of what is known about AWE's application as a class-instruction tool. Specifically, Stevenson's (2016) meta-analysis looked at Criterion and My Access! studies and identified three key constructs in the relevant literature relating to the integration of AWE in the writing classroom which were purpose, action, and use. The purpose of AWE in the classroom was primarily to save teachers time (Chen & Chang, 2008; Li et al., 2015), promote learner autonomy (Chen & Chang, 2008; Grimes & Warschauer, 2010), develop writing processes (El Ebyary & Windeatt, 2010), and raise awareness (Li et al., 2015). Integration of AWE in the classroom can be made possible through scaffolding (Grimes & Warschauer, 2010), embedding in classroom instruction (Li et al., 2015), and assessment and exam preparation (Chen & Chang, 2008).

Caveleri and Dianati (2016) investigate student perceptions of Grammarly in terms of the Technology Acceptance Model (TAM). Overall, students reported that the explanations Grammarly provided helped them understand grammar rules. Grammar books and paper-based exercises on photocopied handouts are portable but lack the direct interactivity with students, which online grammar checkers can provide. Furthermore, feedback from Grammarly "led to reflection about grammar that may not have occurred otherwise" (Caveleri, p. 233). Through a mixed method exploratory design that compared responses from students receiving feedback from Grammarly and students receiving feedback from the teacher, O'Neil and Russell (2019) found that students using Grammarly

responded more positively and enjoyed AWE significantly more than instructor feedback. Students in the Grammarly group were satisfied with the amount of time that was spent on feedback versus the teacher-feedback group. Furthermore, students in the Grammarly group were more likely to state that they had received useful feedback. The researchers recommend that “the program [Grammarly] is used in conjunction with academic learning advisor input as the program is currently not accurate enough for independent use to be justified” (O’Neil & Russel, p. 42). It should be noted that both feedback groups were satisfied with the feedback received but the Grammarly group was significantly happier. O’Neil and Russell (2019) recognized a weakness with AWE related to the inaccuracy of some feedback, and they feel further investigation is needed to identify the errors which Grammarly most frequently missed or misidentified.

The impact of AWE integration on key stakeholders requires a clearer understanding of the context in which the AWE system is used and the content of what is written. To help understand this integration, our study explored the extent that AWE effectively assessed error type, error frequency, and sentence complexity in the EFL context and how that assessment differed across conceptually unique L2 writing genres. This study first identified similarities and differences in error types across L2 writing genres. Next, syntactic complexity and lexical variation were measured. The research questions of the current study were formulated as follows:

1. What are the similarities and differences in error-types among admission test-based writing, descriptive textbook-based essay writing, SNS posts, and SNS comments?
2. What are the similarities and differences in syntactic complexity and lexical variation among admissions test-based writing, textbook-based essay writing, SNS posts, and SNS comments?

### Methods

This corpus analysis study compared error types, error frequency, lexical variation and syntactic variation across four genres of writing - university admission writing tests, textbook-based essays, social media posts, and social media comments. Table 1 displays information on the four sets of corpora.

Table 1  
*Corpora Word Count*

Corpus	<i>M</i>	<i>SD</i>	Words	N
Textbook	2,375	554.03	52,250	22
Test	369	43.87	36,940	100
Posts	711	326.05	29,151	41
Comments	736	431.04	19,136	26

## Corpora

Writing for the test-based corpus came from Yonsei English Learners' Corpus (YELC). Undergraduate students were administered the 2011 Yonsei English Placement Test (YEPT). The YEPT writing section lasted 60 minutes and students completed two questions. Question One called for free writing and Question Two asked students to write an argumentative essay. The YELC classifies composition quality according to the Common European Framework of Reference (CEFR) criteria. All students in the test-based group were between the age of 18 and 19 and were equally parsed according to gender (M = 50, F = 50) and further parsed according to L2 proficiency. Only B1 (N=35), B1+ (N=40), and B2 (N=35) essays were analyzed in order to represent similar proficiency levels to the students who produced the textbook-based essays and SNS-based writing. The following is an example passage taken from a YELC composition:

*To be honest, I think physical punishment should be allowed in schools. In fact besides physical punishment there are other kinds of punishment such as additional homework. But these other types of punishment have not been effective in school...*

Students in the textbook-based writing corpus group followed process writing instruction to complete a series of writing tasks from the book *Writing from Within 2* (Kelly & Gargagliano, 2011). The corpus of textbook-based writing consisted of 52,250 words from 22 English majors attending an L2 writing class. The 22 (M = 4, F = 18) English majors in the textbook writing group were on average 22.1 years old (SD = 1.48). Writing activities involved traditional writing instruction in which the instructor introduced a writing topic, used modeling and writing templates to help students develop ideas, helped students begin first drafts, and discussed methods to improve writing. The six descriptive writing tasks administered had similar difficulty with one another and provided equal amounts of scaffolding via modeling and brainstorming activities. Writing collected came from the first drafts. Task one asked students to describe their favorite place; task two asked students to describe their personal qualities; task three asked students to describe an invention; task four asked students to describe something that changed their life; task five asked students to describe their dream job; task six asked students to describe their personal goals. First drafts from these assignments were compiled for each student and then processed through Grammarly. The following is an example passage from a composition written by a student in the textbook corpus group:

*Minji has a curious personality. She is always interested in working in many different fields. Regarding her most likely personality, a professor would be a great future job for her. She likes to teach, research, and give counsel. First of all, Minji likes teaching. She feels worth when...*

SNS-based compositions were parsed into main posts and comments. At least 300 words were

set as the minimum number of words a student had to produce for inclusion into either the main post or comment corpus. The 41 (M = 18, F = 28) English majors in the Facebook group were on average 22.4 years old (SD = 1.36). Of the 41 students who contributed main posts, only 26 contributed at least 300 words in their comments (Male = 10, Female = 16). Students in the Facebook group were attending a Multimedia English course. According to the International English Language Testing System (IELTS) rubric, students in the textbook and SNS groups had similar L2 proficiency ranging from B1 to B2 CEFR Levels. Examples of Facebook posts and comments are displayed in Table 2. Proficiency levels were assessed by the researcher-instructor embedded with participants during a 16-week semester.

Table 2  
*Examples of Actual Main Posts, Comments and Replies*

**Main Post**

*Hello everyone? How was your weekend?*

*I had a good time. I met my friends at the Igseon-dong station. There were so many people. There stores are great. It's very unique. And, it is a single-story building. The interior was good. At first, we tried to eat dumplings. But...(Continued)*

**Comment 1**

(reflective) *Wow ~ It was so delicious! I will go there!*

**Reply**

*Yes, it will be good!*

**Comment 2**

(substantive) *Oh! I went to Igseon-dong last week!! The alley was so beautiful. I saw the restaurant! But I didn't go. There were many people waiting! I went to another restaurant! I ate spaghetti!!*

*There was no restaurant sign. I want to go to this place again.*

## Procedures

First, the text from each corpus needed prepared for analysis. Methods for preparing documents differed among corpora because of varying word count among group members. For the textbook and SNS groups, each students' writing was analyzed separately from other students in their corpus group. In the textbook group, all of the writing tasks from the same student were combined and then analyzed so there was a total of 22 Grammarly reports produced. Similar steps were taken for the SNS post and SNS comment group creating 41 and 26 Grammarly reports respectively. For the test-based writing group, students were aggregated into groups of five and Grammarly reports were created for each of these groups, which resulted in a total of 20 Grammarly reports.

Grammarly performance reports provided analytics on the error-types, syntactic complexity, and overall writing scores. Grammarly offers free automatic feedback on critical errors while the commercial version provides feedback on advanced errors and allows users to download analytics (i.e., performance reports). The commercial version was used for this study. Figures 1 and 2 display samples of a performance report.

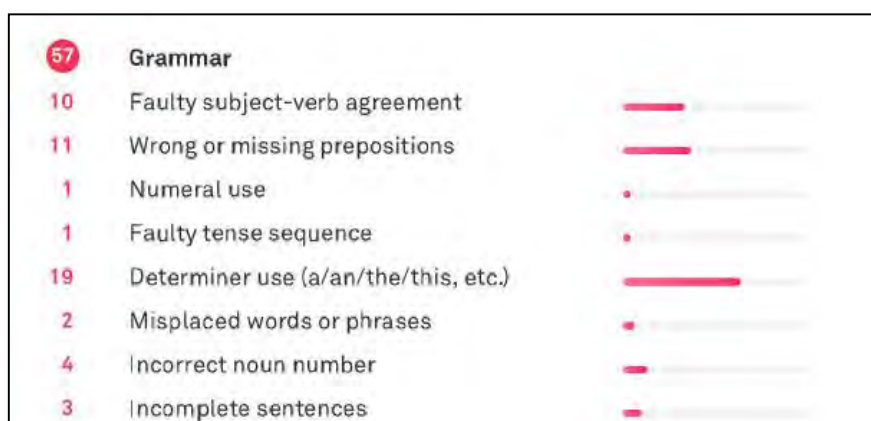
Figure 1

*Example of Grammarly Performance Report Part 1*



Figure 2

*Example of Grammarly Performance Report Part 2*

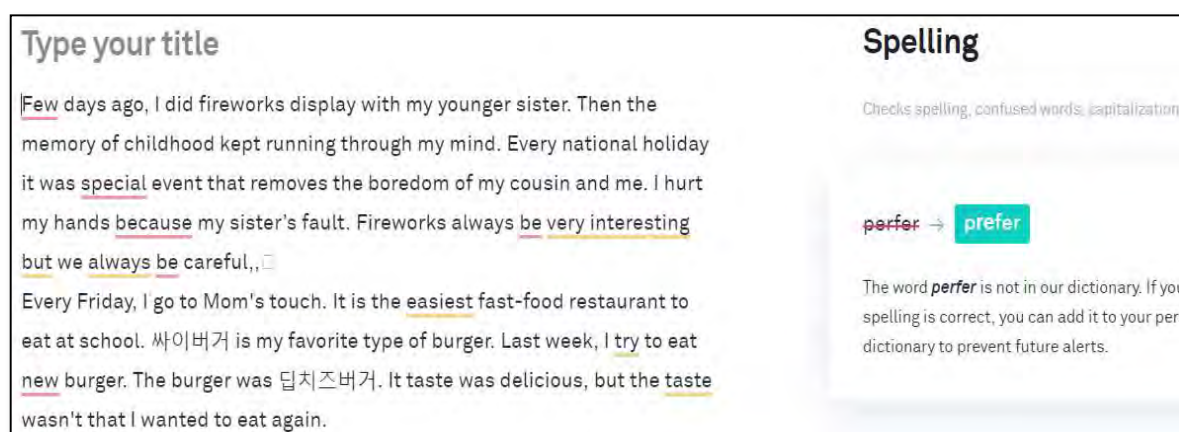


When using Grammarly, the composition is displayed on the left side of the screen and corrections can be made in real time. Figure 3 shows an example of the text editor panel for

Grammarly. This information is provided to the user in real time so that corrections can be applied while composing text.

Figure 3

## Example of the Grammarly Text Editor



While Grammarly identifies over 300 error types, this study only analyzed error-types that occurred 0.3 times out of a hundred words for at least one of the three writing sets. Table 3 displays error categories analyzed by Grammarly along with brief descriptions and examples of some error-types. Errors that occurred less frequently were aggregated into error groups (e.g., punctuation, grammar, conventions).

Table 3  
*Error-Type Categories*

Error Type	Description	Feedback Given by Grammarly
<b>Punctuation</b>	Corrects missing and misused punctuation	It appears that you are missing a comma before the coordinating conjunction <b>but</b> in a compound sentence.
Complex or Compound Sentence	<i>e.g., Run on sentences</i>	
Comma in Clause	<i>Well [,] I am not sure about that</i>	
<b>Grammar</b>		
Determiner	<i>I think [the] bottle is so cute.</i>	
Incomplete Sentence	<i>Never on time.</i>	
Subject-Verb	<i>I drove [rode] my bike to school.</i>	
Preposition Error	<i>He goes on his work.</i>	
<b>Variety</b>	Provides synonyms for repeated words	The word <b>order</b> appears repeatedly in this text. Consider using a synonym in its place.
<b>Convention</b>	Checks spacing, capitalization, and dialect-specific spelling	The numeral <b>2</b> is used instead of the word spelled out. Consider spelling out the number.
<b>Spelling</b>	Corrects misspellings, misused words, and capitalization	
<b>Conciseness</b>	Eliminates wordiness and redundancy	It appears that <b>Actually</b> , may be unnecessary in this sentence.

In addition to error-types, certain properties of syntactic and lexical diversity were analyzed.



Grammarly produced syntactic analytics for word and sentence length. Furthermore, unique words and rare words were also identified and compared among writing sets. Lu's (2012) lexical complexity analyzer was used to identify noun, verb, and modifier variation. Lexical sophistication measures the proportion of relatively unusual or advanced words in the learner's text (Read, 2000), and lexical variation (i.e., lexical diversity) refers to the range of a learner's vocabulary as displayed in their language use. Table 4 displays the indices and the formulas used to calculate values.

Table 4  
*Lexical and Syntactic Indices*

Word Length	Total words / Total Characters
Sentence Length	Total Sentences / Total Words
Noun Variation*	Different Nouns / Total Nouns
Verb Variation*	Different Verbs / Total Verbs
Modifier Variation*	Different Modifiers (adjective + adverbs) / Total Modifiers
Unique Words	Measures vocabulary diversity by calculating the percentage of words used only once in your document
Rare Words	Measures depth of vocabulary by identifying words that are not among the 5,000 most common English words.

Note: \*Variation indices were computed using Lu (2013) lexical complexity analyzer.

## Data Analysis

The software package SPSS 24.0 was used to analyze data. Analysis of variance (ANOVA) is a common statistical procedure used to determine whether any statistically significant difference between the means of two or more unrelated groups. For research question one, mean score comparison with one-way ANOVA was carried out to identify statistically significant differences in error types among test-based, textbook-based, and SNS-based corpora. For research question two, a similar mean score comparison with one-way ANOVA was used to identify statistically significant differences for lexical and syntactic complexity among the writing sets.

## Results and Discussion

Answering research question one begins by describing the mean scores for error types recognized by Grammarly (Table 5). Details are given for the most frequent errors and statistically significant differences in error types between genres. Research question two goes on to describe the lexical and syntactic similarities and differences across genres.

Table 5  
Descriptive Statistics for Error Categories

Error Category		<i>M</i>	<i>SD</i>	Error Category		<i>M</i>	<i>SD</i>
<b>Errors Per 100 Words</b>	1	8.80	3.24	<b>Variety<sup>a</sup></b>	1	0.63	0.27
	2	13.13	2.37		2	1.24	0.44
	3	10.85	4.01		3	0.69	0.51
	4	11.83	3.14		4	0.48	0.42
	Total	11.05	3.56		Total	0.73	0.50
<b>Grammar</b>	1	3.12	1.40	<b>Conventions</b>	1	1.01	1.03
	2	4.81	1.17		2	0.72	0.78
	3	2.71	1.34		3	0.52	0.68
	4	3.28	1.31		4	1.09	1.79
	Total	3.32	1.50		Total	0.79	1.14
Determiner	1	1.36	0.65	<b>punctuation</b>	1	1.02	0.46
	2	2.68	0.61		2	0.99	0.22
	3	1.34	0.83		3	1.13	0.80
	4	1.45	0.75		4	2.29	1.22
	Total	1.62	0.89		Total	1.36	0.95
Incomplete Sentences	1	0.25	0.19	Comma in Clause	1	0.34	0.22
	2	0.24	0.12		2	0.42	0.17
	3	0.37	0.35		3	0.30	0.41
	4	0.83	0.67		4	1.04	0.90
	Total	0.43	0.46		Total	0.51	0.59
Faulty Subject-Verb Agreement	1	0.32	0.39	Punctuation in Compound/Complex Sentence	1	0.51	0.29
	2	0.57	0.32		2	0.48	0.19
	3	0.17	0.20		3	0.59	0.42
	4	0.18	0.24		4	0.65	0.53
	Total	0.27	0.32		Total	0.57	0.39
Preposition	1	0.45	0.19	<b>Spelling</b>	1	1.46	0.88
	2	0.46	0.15		2	3.06	1.04
	3	0.26	0.22		3	1.96	1.76
	4	0.30	0.31		4	2.34	1.85
	Total	0.35	0.24		Total	2.15	1.60
<b>Conciseness (Wordiness)</b>	1	0.33	0.18				
	2	0.24	0.19				
	3	0.54	0.46				
	4	1.16	0.74				
	Total	0.59	0.58				

Note: Textbook-based = 1, Test-Based = 2, SNS Posts = 3, SNS Comments = 4; Variety<sup>a</sup> refers to repeated word choice errors.

## Grammar Errors

Results from answering research question one showed students committed grammar mistakes the most out of the Grammarly error-type groups. Of the four corpora, students in the test-taking condition produced the most grammar mistakes followed by Facebook posts and comments and then text-book based writing. As indicated in Table 6, these differences were statistically significant. Students committed more mistakes under time-constrained testing conditions.

Among error types in the grammar category, determiner errors were the most common for all groups, with the test-based writing conditions resulting in twice as many determiner errors than the other groups. Determiners indicate whether or not a noun is definite or indefinite and are common

among South Korea L2 writers (Lee, Chodorow, & Gentile, 2016). The most common determiner errors related to articles, which are difficult for L2 English learners (Han, Martin, & Leacock, 2006). One of the most difficult challenges for L2 writers is mastering the use of English articles (Han et al., 2006). Correctly using definite and indefinite articles can be difficult in different linguistic environments and even with the inclusion of instructor-provided corrective feedback (Ferris, Chaney, Komura, Roberts, & McKee, 2000). Grammarly drastically reduces article and other determiner errors.

There were a number of other grammar error-types that differed in ratio among the writing sets. Facebook comments had three times as many incomplete sentence errors than either the test or textbook-based writing groups and twice as many as the Facebook main posts. Incomplete sentences are common in SNS communication because ideas can be *completed* with non-textual multimodal (e.g., photos, links, or video) communication (Bailey et al., 2017). Students write shorter sentences in Facebook comments, and these sentences can often be incomplete statements referencing a point of discussion located further up the discussion chain, or in comments that are expressing an emotion, and therefore more incomplete sentence error and incomplete comparison errors occur.

Table 6  
*Bonferroni Post Hoc Analysis of Error Categories across Genre*

Error Type			<i>MD</i>	<i>SE</i>	<i>p</i>
<b>Grammar</b>		2 1	1.69	0.41	.000
		3	2.10	0.36	.000
		4	1.53	0.39	.001
Determiner		2 1	1.32	0.23	.000
		3	1.33	0.20	.000
		4	1.23	0.22	.000
Incomplete Sentence		1 4	-0.58	0.12	.000
		2 4	-0.59	0.12	.000
		3 4	-0.46	0.10	.000
Incorrect Noun		2 3	0.11	0.04	.017
Faulty Subject-Verb Agreement		2 1	0.24	0.09	.038
		3	0.40	0.08	.000
		4	0.39	0.08	.000
Preposition		3 1	-0.18	0.06	.017
		2	-0.20	0.06	.011
<b>Conciseness</b>		4 1	0.83	0.14	.000
		2	0.92	0.14	.000
		3	0.62	0.12	.000
<b>Variety</b>		2 1	0.61	0.14	.000
		3	0.54	0.12	.000
		4	0.76	0.13	.000
<b>Spelling</b>		1 2	-1.59	0.47	.006
		4	-0.88	0.44	.304
<b>Punctuation</b>		4 1	1.27	0.23	.000
		2	1.29	0.24	.000
		3	1.15	0.20	.000
Comma in Clause		4 1	0.70	0.15	.000
		2	0.62	0.15	.001

Note: Textbook-based = 1, Test-Based = 2, SNS Posts = 3, SNS Comments = 4

Incorrect noun choice was the next error-type in the grammar category that revealed a significant difference across groups. Test-based writing resulted in more noun choice mistakes than Facebook posts. Test-based writing consisted of longer and rarer words than Facebook writing. A similar pattern of more errors occurring under test-based conditions appears with subject-verb agreement (e.g., He *have* to go – He *has* to go) and prepositions (e.g., *at* Monday – *on* Monday). The students writing essays committed more grammar mistakes whether under the time time-dependent test-based condition or time-independent textbook-based condition, and this was attributed to the inclusion of more rare words and longer sentences.

AWE software like Grammarly can be used to address what Ferris (1999) termed *treatable* errors such as verb tense and form, subject-verb agreement, article usage, plural and possessive noun endings, and sentence fragments, while writing instructors can provide an informed complementary feedback approach for identifying errors within a structural or textual whole (Hamilton, 2015).

Incomplete reconstruals are common errors committed by East Asian writers in countries like

China (Liardet, 2015). Grammarly was able to consistently identify incomplete reconstructions like *the understand between people* and correct them with feedback like *the understanding between people*.

### Conciseness, Convention, and Spelling Errors

Error types concerning conciseness, or wordiness, also revealed differences in error ratios across groups. Facebook posts resulted in twice as many conciseness errors and Facebook comments resulted in four times as many than textbook-based and test-based writing. The reasoning for this was attributed to casual language which is common in Facebook posts. Weak adjectives such as “like”, “really”, and “very” were frequently found in Facebook writing. To compensate for the overuse of *weak* words, Grammarly provided more sophisticated alternatives.

Test-based and Facebook posts revealed more writing in the passive voice. While not incorrect, the passive voice is often considered a bad writing habit. Active voice is preferred because it makes writing stronger and more direct. Writing from test-taking conditions resulted in twice as many variety, or word choice, errors. For example, if the student wrote about smoking in public then the Grammarly program would offer alternatives to “smoking” due to the overuse of the word. Conciseness, passive voice, and word-choice errors are not technically mistakes, but characteristics of writing quality that Grammarly attempts to improve.

Convention errors (e.g., capitalization and spacing errors) occurred in all sets of writing. Facebook posts produced the least amount of convention errors but not at a statistically significant level. With Facebook, the higher rate of accuracy may be attributed to asynchronous writing and access to devices that correct automatically. Spelling errors (spelling and capitalization of proper nouns) were the second most frequently occurring error type category recognized by Grammarly. Textbook-based writing resulted in the least amount of spelling mistakes, while Facebook posts and test-based writing produced the most. Many of the words misspelled in the test-based group were high-frequency words (i.e., most common 1000 words) such as *thought* (thought), *frend* (friend), and *raining* (raining) and were attributed to the time-dependent nature of the task. Furthermore, students in the test-based condition had no access to spell-checkers or dictionaries which likely contributed to these convention and spelling errors. In addition to misspelled words, Facebook posts revealed numerous examples of proper nouns starting with lowercase letters.

The four writing sets exhibited high levels of spelling errors which is also common among L2 writers (Lee et al., 2016). SNS writing had more spelling mistakes related to unknown words, which is expected with greater topic variety and inclusion of Korean names (*bulgogi*, meaning meat) and locations (*Mapo Naru*, a popular restaurant in Seoul). Instructors are recommended to provide examples of when to avoid recommendations from Grammarly when country-specific terms such as unknown proper nouns like names and locations are used by the writer.

## Punctuation Errors

Punctuation errors were the third most common error category across the four corpora. Facebook comments resulted in twice as many punctuation errors as the other writing sets and this was attributed to mistakes with comma usage in clauses. Facebook comments use ellipsis to signify pause and emphasize an emotion not adequately captured by an exclamation point. For instance, in comments, students would write, “Oh my... how cute!” Here, we see an utterance expressing engagement “Oh my...” followed by excitement as indicated by “how cute!” These are examples of a SNS genre-specific writing styles and not necessarily writing errors.

Test-based and textbook-based writing was more complex than spoken language (Halliday, 1989), and this is reflected in the genre of SNS writing. Facebook writing was grammatically less complex than test-based and textbook-based writing. Incomplete sentence errors were more frequent in Facebook. The use of Facebook benefits language learners and creates communities to practice language learning in informal environments (Newgarden, 2009) which results in the use of casual languages in the form of utterances (Bailey et al., 2017). Frequent use of utterances in writing resulted in shorter sentences and greater frequency of punctuation errors related to incomplete sentences.

Writing that occurred on Facebook had more punctuation mistakes. These included punctuation mistakes with sentence-endings and punctuation mistakes within clauses. Communication with SNS is similar to the utterances that occur during face to face communication than writing that occurs through textbook-based essay writing and therefore Facebook writing was more casual than writing emanating from process writing instruction. This informal nature with Facebook writing allows for more freedom with punctuation use. Examples of loose punctuation rules applying to Facebook writing include triple ellipses [...] inside sentences to reference pauses, and double exclamation points at the end of sentences to emphasize emotion.

Research question two investigated the similarities and differences in lexical and syntactic diversity across the four sets of writing. Table 7 displays mean scores and Table 8 shows results from Bonferroni posthoc analysis of the measured indices. Statistically significant differences were recognized in each of the measured syntactic and lexical indices. Students in the textbook-based group wrote longer sentences and words while students in the SNS-based group used a wider variety of nouns, verbs, and modifiers (e.g., adjectives and adverbs).

Table 7  
*Lexical and Syntactic Indices*

Indices	Corpus	M	SD	Indices	Corpus	M	SD
Different Words/Random 50 Words	1	37.2	1.32	Adverb Variation	1	.05	.007
	2	37.1	1.32		2	.05	.007
	3	37.3	1.67		3	.07	.022
	4	37.1	1.78		4	.08	.041
Verb Variation	1	.10	.01	Modifier Variation	1	.13	.015
	2	.10	.01		2	.13	.011
	3	.14	.04		3	.17	.034
	4	.16	.07		4	.19	.073
Noun Variation	1	.46	.05	Word Length	1	4.3	0.09
	2	.42	.04		2	4.6	0.25
	3	.61	.08		3	4.1	0.19
	4	.69	.15		4	3.8	0.79
Adjective Variation	1	.08	.01	Sentence Length	1	10.8	1.32
	2	.08	.01		2	11.9	1.17
	3	.10	.02		3	9.3	1.91
	4	.11	.04		4	6.8	1.02

Note: Textbook-Based = 1, Test-Based = 2, SNS Posts = 3, SNS Comments = 4

Table 8  
*Bonferroni Post Hoc Analysis of Lexical and Syntactic Complexity Indices*

			MD	SE	p				MD	SE	p
Verb Variation	1	3	.037	.010	.003*	Modifier Variation	1	3	.041	.011	.002
		4	.057	.011	.000**			4	.062	.012	.000**
	2	3	.036	.011	.006*		2	3	.041	.012	.004*
		4	.056	.012	.000**			4	.062	.013	.000**
Noun Variation	1	3	.153	.025	.000**	Word Length	1	4	0.538	0.121	.000**
		4	.237	.027	.000**		2	3	0.497	0.114	.000**
	2	3	.194	.026	.000**			4	0.813	0.124	.000**
		4	.278	.028	.000**						
Adjective Variation	3	4	.084	.024	.004*	Sentence Length	1	3	1.487	0.431	.005*
	1	4	.024	.007	.005*			4	4.258	0.472	.000**
	2	4	.025	.007	.005*		2	3	2.640	0.444	.000**
	1	3	.022	.007	.006*			4	5.411	0.485	.000**
Adverb Variation							4	3	-2.771	0.409	.000**
		4	.037	.007	.000**						
	2	3	.023	.007	.005*						
		4	.038	.007	.000**						

Note: Textbook-Based = 1, Test-Based = 2, SNS Posts = 3, SNS Comments = 4; Bonferroni adjustment, .05/8 = .0063\*, .01/8 = .0013\*\*

Results from answering research question two explored similarities and differences in syntactic and lexical complexity among the different writing genres. Facebook posts and comments resulted in fewer instances of awkward sentences and misused words and this was attributed to shorter

2020 Volume 15 Issue 2 2020 ISSN 2094-3938

sentence length and simpler vocabulary choice associated with the genre of SNS writing.

Topics for Facebook discussion are near limitless; suggesting a wider variety of vocabulary would be practiced. Even greater differences may exist between test-based writing and weekly class assignment writing tasks (e.g., textbook-based or SNS-based writing). The test-based corpus in the current study came from university entrance exam questions. Academic stakes are greater with test-based writing, and this may result in differences in writing complexity which could lead to errors related to sentence and word complexity as more advanced structures are attempted.

### **Pedagogical Implications**

Certain L2 writing mistakes were constant throughout each genre but unique differences did exist and this leads us to make targeted pedagogical recommendations. Students in the test-based corpus group made considerably more grammar and wrong-word choice mistakes than in the textbook-based and SNS-based groups, possibly due to lack of time afforded through asynchronous writing. To help students practice for high stakes writing tests, we recommend students keep an archive of automatically generated feedback so they can practice metacognitive strategies, like reflection, to decrease common errors over time. Clarity issues with high frequency errors must be achieved before test raters can consider factors related to vocabulary and sentence complexity. Overcoming high frequency mistakes with article usage can be achieved with consistent corrective feedback. The test-based group committed twice as many determiner errors related to article usage so this is an area that should be addressed early with test preparation programs that incorporate AWE.

The textbook-based group fell in between test-based and SNS-based writing for many of the Grammarly metrics observed. The extra time and process writing instruction resulted in relatively equal sentence complexity with test-based writing but fewer errors. Low word variation echoed what was observed with the test-based group and this was partly attributed to students working off of textbook writing models. The same students borrowed language from the same textbook examples just as the same students in the testing condition borrowed language from the essay test questions. Contrarily, students in the SNS-based group were responsible for developing their own writing topics, and this resulted in a broader range of vocabulary use. We recommend students use automatic writing aides like Grammarly to identify overly used words and replace them with synonyms. Non-native English-speaking university students require paraphrasing skills. Several scholars (Leask, 2006; Liu, 2005) have raised concerns of plagiarism among different cultural groups highlighting a need for paraphrasing skills among L2 English speaking university students. Novice writers rely on source text language extensively while more advanced ones are capable of paraphrasing (Keck, 2014). Practicing paraphrasing skills early on with one's own writing may better equip students for paraphrasing in academic or business writing. A final recommendation for the use of AWE with process writing instruction pertains to the timing of AWE implementation. Grammarly may be most valuable for L2 writers immediately after composing first drafts. With immediate feedback, students can recognize possible areas of concern before sending their writing to the instructor.



The SNS-based writing corpus revealed stark differences with test-based and textbook-based writing, especially concerning word variation, sentence complexity, and error frequency. SNS-based writing emanating from the corpus analyzed in the current study can best be likened to shared class diaries of daily activities and personal interests (Bailey et al., 2017). The increased word variation is ascribed to the idiosyncrasies among the Facebook group members' interests and daily events. Students contributing to SNS discussion forums are left to their own free will to choose what topics will be discussed, and this created real world opportunity to use a wide range of vocabulary. Word variation and idea-generation was further amplified with the use of multi-modal communication common in online discussion threads. Over 95 percent of Facebook posts were in reference to images, and many of the images were accompanied by hyperlinks, emoticons, or videos.

Instructors are recommended to host SNS forums for students learning to use new vocabulary in authentic settings. The SNS group used simpler sentence structure and vocabulary than the other groups, and this was especially true for SNS comments and replies. The lower stakes involved with SNS writing meant that the students had no need to impress their instructors or test raters. The SBS activity was graded as pass/fail with main criteria being word count and the number of contributions (i.e., posts and comments). Students had nothing to gain by using more sophisticated vocabulary or sentences and this resulted in fewer accuracy issues and consequently greater clarity. Students were sharing their writing with classmates so they may have invested more energy on accuracy and clarity to avoid looking foolish (Krashen, 1985).

Students in the SNS group had model examples of good writing. Role model examples of good posts were provided by L2 proficient writers and this acted as samples for less capable writers to follow. An advanced writer may write, "Today I went to the *beach* with my *friends* and afterward we had *seafood* for *dinner*." A less L2 proficient classmate can use this sentence as a template to write something like, "This *morning* I went to *church* with my *family* and afterward we had *pizza* for *lunch*." Bailey et al. (2017) recognized that students were borrowing language from one another and at the same time practicing new vocabulary in a real world context. Peer modeling is a characteristic of collaborative learning that increases self-efficacy beliefs (Bandura, 1996), and role model posts on SNS helped others produce compositions with fewer accuracy mistakes. In addition to borrowing language from their classmates' posts, students could practice recursive writing strategies (e.g., second drafts) by borrowing language from their own past posts. A student can use similar language to describe both a dinner with friends and a lunch with family. After a few posts with similar communication goals, the accuracy improves because students are following similar sentence patterns and using similar adjectives but for different communication purposes.

AWE implementation is not a replacement for traditional writing strategies but instead an addition to existing ones. Grammarly should be added to the language learner's L2 writing strategy repertoire.

Writing strategies occur at different stages of the writing process. We recommend Grammarly's use during the when-writing and post-writing stages. Planning, brainstorming, and

outlining should be carried out without concern of writing accuracy. L2 students are even encouraged to use their L1 during the pre-writing stage for idea-generation. Ericsson (2006) recommends automatically generated feedback for writing assistance rather than writing assessment, and the strategy of using AWE as a personal writing assistant is what makes platforms like Grammarly so popular. Successful language learners use a wider array of language learning strategies than their less successful counterparts (Oxford, 1990), suggesting AWE be an addition to the learner's strategy toolbox.

### Conclusion

Results from this study provide insight into the types of errors Grammarly recognizes among conceptually unique genres of L2 writing. Instructors of English as a second or foreign language can apply findings in the present study to support the development of their students' writing skills. Students using AWE can save time and increase their confidence when writing in a second language because of fewer accuracy mistakes, and instructors can focus their corrective feedback on higher order writing issues related to rhetoric, tone, cohesion, and style.

Several themes were recognized among error types, error frequency, and writing complexity for the test-based, textbook-based, and SNS-based corpora. Grammarly was successful at identifying local level errors in L2 writing regardless of the writing genre. However, distinct differences were recognized. High stakes testing resulted in more risk taking with vocabulary and sentence complexity which came at the cost of readability (i.e., clarity). Variation errors are mistakes due to repeating the same words frequently. Such vocabulary repetition was common with test-based writing, indicating a need to use AWE programs like Grammarly for paraphrasing practice. Grammarly was most successful at identifying and correcting determiner errors which proves valuable for East Asian students because such errors are common with Chinese, Korean, and Japanese L2 English writers.

Further errors recognized were related to spelling, punctuation, and wordiness. In general, Grammarly incrementally improved writing compositions. A low- quality composition cannot be transitioned into a high quality of writing with Grammarly alone, but instead writing quality improves at incremental levels.

A few limitations with this study are worth mentioning. Findings may not generalize outside the Sino-Tibetan-Austronesian family of languages. Writing samples were collected from intermediate L2 proficiency students. Error types and error frequency would be different depending on the writers' L2 proficiency levels (Liardét, 2015). Finally, results may be different depending on which AWE programs is being used. For instance, Criterion ([www.criterion.ets.org](http://www.criterion.ets.org)) and Virtual Writing Tutor ([www.virtualwritingtutor.com](http://www.virtualwritingtutor.com)) are more focused on giving feedback to students in an academic setting while Grammarly targets users who are both L1 and L2 students or working professionals.

Future AWE research should explore the effect of integrating tools like Grammarly into the

L2 writing process. AWE technology should not replace existing L1 or L2 writing strategies but instead complement them. Future study may wish to investigate how AWE platforms can work in concert with one another.

Moving forward, EFL/ESL educators should consider how best to implement AWE technology into their writing and communication programs. There is no substitute for human feedback when it comes to the nuances of global level writing errors related to meaning-making and cohesion but such mistakes cannot be addressed when instructors are overwhelmed with local level, treatable, errors.

AWE tools like Grammarly increase writing fluency by saving time composing ideas and this equates to greater language output but at what cost? The ultimate goal should be to help the writer become self-reliant. There is a need now to better understand how dependency on writing aid technology influences self-reliance.

## References

- Anson, C. (2006). Can't touch this: Reflections on the servitude of computers as readers. In P. F. Ericsson, R. Haswell (eds), *Machine scoring of human essays* (pp. 38–56). Logan, UT: Utah State University Press.
- Bailey, D. B., Park, I.-W., Haji, S. A. (2017). An investigation of Facebook for language learning: Better understanding perceptions and participation. *CALL-EJ*, 18(2), 12-28. Retrieved from <http://callej.org/journal/18-2/Bailey-Park-Haji2017.pdf>
- Burston, J. (2008). Review of BonPatron: An online spelling, grammar, and expression checker. *CALICO Journal*, 25(2), 337-347. doi:10.1558/cj.v25i2.337-347
- Cavaleri, M., & Dianati, S. (2016). You want me to check your grammar again? The usefulness of an online grammar checker as perceived by students. *Journal of Academic Language and Learning*, 10(1), 223-236. Retrieved from <https://journal.aall.org.au/index.php/jall/issue/view/22>
- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 12(3), 267-296. [https://doi.org/10.1016/S1060-3743\(03\)00038-9](https://doi.org/10.1016/S1060-3743(03)00038-9)
- Chen, C.-F., & Cheng, W.-Y. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning and Technology*, 12(2), 94-112. Retrieved from <https://www.lltjournal.org/item/2631>
- Chung, G. K. W. K., & Baker, E. L. (2003). Issues in the reliability and validity of automated scoring of constructed responses. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 23-40). Mahwah, NJ: Lawrence Erlbaum.
- Dikli, S. (2010). The nature of automated essay feedback. *CALICO Journal*, 28(1), 99-134. Retrieved from <https://www.jstor.org/stable/calicojournal.28.1.99>
- El Ebyary, K., & Windeatt, S. (2010). The impact of computer-based feedback on students' written work. *International Journal of English Studies*, 10(2), 121-142. Retrieved from <https://files.eric.ed.gov/fulltext/EJ936915.pdf>
- Ericsson, P. F. (2006). The meaning of meaning: Is a paragraph more than an equation. In P. F. Ericsson & R. H. Haswell (Eds.), *Machine scoring of student essays* (pp. 28-37). Logan, UT: Utah State University Press.
- Ferris, D. R. (2004). The “grammar correction” debate in L2 writing: Where are we, and where do we go from here? *Journal of Second Language Writing*, 13(1), 49-62. <https://doi.org/10.1016/j.jslw.2004.04.005>
- Ferris, D. R., Chaney, S. J., Komura, K., Roberts, B. J., & McKee, S. (2000). Perspectives, problems, and practices in treating written error. In Colloquium presented at International TESOL Convention, Vancouver, B.C., March 14-18, 2000.
- Gauthier, M. (2013). Anglophone high school boys' engagement and achievement in editing their French writing using the Bon Patron Pro. *Journal of Classroom Research in Literacy*, 6, 24-35. Retrieved from <https://jcr.l.library.utoronto.ca/index.php/jcr/article/view/16093>

- Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *Journal of Technology, Language, and Assessment*, 8(6), 1-43. Retrieved from <https://ejournals.bc.edu/index.php/jtla/article/view/1625>
- Halliday, M. A. K. (1989). *Spoken and Written Language*. Oxford: Oxford University Press.
- Hamilton, C. E. (2015). The contributions of systematic functional grammar to the error analysis framework. *TESOL International Journal*, 10(1), 11-28. Retrieved from <http://tesol-international-journal.com/112/volume-10-issue-1-2015/>
- Han, N.-R., Chodorow, M., & Leacock, C. (2006). Detecting errors in English article usage by non-native speakers. *Natural Language Engineering*, 12(2), 115-129. doi:10.1017/S1351324906004190
- Hoang, G. T. L., & Kunnan, A. J. (2016). Automated essay evaluation for English language learners: A case study of MY Access. *Language Assessment Quarterly*, 13(4), 359-376. <https://doi.org/10.1080/15434303.2016.1230121>
- Katayama, A. (2007). Japanese EFL students preferences towards correction of classroom oral errors. *The Asian EFL Journal*, 9(4), 289-305. Retrieved from <https://www.asian-efl-journal.com/main-journals/japanese-efl-students-preferences-toward-correction-of-classroom-oral-errors/>
- Keck, C. (2014). Copying, paraphrasing, and academic writing development: A re-examination of L1 and L2 summarization practice. *Journal of Second Language Writing*, 25, 4-22. <https://doi.org/10.1016/j.jslw.2014.05.005>
- Kelly, C., & Gargagliano, A. (2011). *Writing From Within 2*. Cambridge, UK: Cambridge University Press.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. New York: Longman.
- Leask, B. (2006). Plagiarism, cultural diversity and metaphor—implications for academic staff development. *Assessment & Evaluation in Higher Education*, 31, 183-199. <https://doi.org/10.1080/02602930500262486>
- Liardét, C. (2015). Academic literacy and grammatical metaphor: Mapping development. *TESOL International Journal*, 10(1), 29-46. Retrieved from <http://tesol-international-journal.com/112/volume-10-issue-1-2015/>
- Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, 27(1), 1-18. <https://doi.org/10.1016/j.jslw.2014.10.004>
- Liao, H. (2016). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach. *System*, 62(1), 77-92. <https://doi.org/10.1016/j.system.2016.02.007>
- Liu, D. (2005). Plagiarism in ESOL students: Is cultural conditioning truly the culprit? *ELT Journal*, 59, 234-243. <https://doi.org/10.1093/elt/cci043>
- Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, 96(2), 190-208. [https://doi.org/10.1111/j.1540-4781.2011.01232\\_1.x](https://doi.org/10.1111/j.1540-4781.2011.01232_1.x)

- McGee, T. (2006). Taking a spin on the Intelligent Essay Assessor. In P. F. Ericsson & R. H. Haswell (Eds.), *Machine scoring of student essays* (pp. 79-92). Logan, UT: Utah State University Press.
- Nadasdi, T., & Sinclair, S. (2007). Anything I can do, CPU can do better: A comparison of human and computer grammar correction for L2 writing using BonPatron.com. Retrieved from <http://bonpatron.com>
- Newgarden, K. (2009). Annotated bibliography: Twitter, social networking and communities of practice. *TESL-EJ*, 13(2), 2-13. Retrieved from <http://www.tesl-ej.org/pdf/ej50/m2.pdf>
- O'Neill, R., & Russell, A. M. T. (2019). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology*, 35(1), 42-56. <https://doi.org/10.14742/ajet.3795>
- Potter, R., & Fuller, D. (2008). My new teaching partner? Using the grammar checker in writing instruction. *English Journal*, 98(1), 36-41. Retrieved from <https://www.jstor.org/stable/40503205?seq=1>
- Read, J. (2000). *Assessing vocabulary*. Cambridge, UK: Cambridge UP.
- Stevenson, M. (2016). A critical interpretive synthesis: The integration of automated writing evaluation into classroom writing instruction. *Computers and Composition*, 42(1), 1-16. <https://doi.org/10.1016/j.compcom.2016.05.001>
- Stevenson, M., & Phakiti, A. (2014). The effect of computer-generated feedback on the quality of writing. *Assessing Writing*, 19(1), 51-65. <https://doi.org/10.1016/j.asw.2013.11.007>
- Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Eds.), *Sociocultural theory and second language learning* (pp. 97-114). London: Oxford University Press.
- Ware, P. (2011). Computer-generated feedback on student writing. *Tesol Quarterly*, 45(4), 769-774. <https://doi.org/10.5054/tq.2010.272524>
- Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36(1), 90-102. <https://doi.org/10.1016/j.asw.2018.02.004>

## Teaching Reporting Verbs to English as a Second Language Undergraduate Writers in the Academic Context

Zhao Jun\*

*Augusta University*

### Abstract

This article reports on a pedagogical trial of teaching reporting verbs for source integration in a North American ESL (English as a second language) composition class. Source incorporation challenges many novice ESL writers in the academic context as they are unable to view writing as an interactional activity, which requires proper presentation of different sources. This problem is often manifested in their monotonous, ineffective choice of reporting verbs. Advancing Halliday's Systemic Functional Linguistics (1978, 1994), Martin and White's Appraisal Theory (2005) further studies Interpersonal metafunction by focusing on speaker/writer attitude/evaluation in language use. It serves as the theoretical framework in this pedagogical trial which then helps student writers understand how reporting verbs could indicate different stances and create interactions in writing. This two-week pedagogical trial followed the teaching cycle of modelling, joint construction and independent construction recommended in the genre approach to help students to better deploy reporting verbs. Their writing samples before and after this pedagogical trial were collected, together with a final survey of their perceptions of the teaching activities, as evidence for the (in)efficiency of this pedagogical trial. Analysis of students' original and revised writings revealed progress but some lingering problems in their reporting verbs. Survey data indicated their enhanced understanding of why and how to integrate sources via reporting verbs, and their positive reactions to the teaching activities. Finally, the implications of this pedagogical intervention are discussed in the academic context for writing.

**Key Words:** reporting verb; source integration; Interpersonal metafunction; Appraisal; genre approach; English as a second language writing

---

\*[juzhao@augusta.edu](mailto:juzhao@augusta.edu)

## Introduction

In academic writing, creating interactions between sources and writer's ideas is extremely important, as writers need to engage in conversations with others to advance knowledge (Kwon, Staples & Partridge, 2018; Lee,

Hitchcock & Casal (2018). However challenging, this is an essential writing skill to be developed. Second language (L2) writing studies on source incorporation have revealed students' general incompetence in this area (Charles, 2006a; Thompson & Ye, 1991; Wingate, 2012). One useful linguistic tool to create interaction in writings that utilize sources is reporting verbs (Bloch, 2010; Charles, 2006a, 2006b; Kwon, Staple & Partridge, 2018; Lee, Hitchcock & Casal, 2018; Thompson & Ye, 1991). Here, forms and semantic categories of reporting verbs (Charles, 2006a, 2006b; Thompson & Ye, 1991) have been widely studied, but the stance of reporting verbs (Coffin, 2009; Thompson & Ye, 1991) has not received much-deserved attention yet.

Halliday's Systemic Functional Linguistics (1978, 1994) (hereafter, SFL) explores the quintessential nature of language. Its Interpersonal Metafunction considers language as a process of "social interaction", a "mode of doing" and a "way of acting" (Halliday & Hasan, 1985:17). Exploring the interpersonal aspect of language use, Appraisal Theory (Martin & White, 2005) furthers SFL and specifically investigates how entities are evaluated in terms of Attitude, Engagement and Graduation. These could be great resources to guide and assist students to understand the interactive nature of writing. However, the pedagogical application of SFL in the language-teaching field is scarce (McCabe, Gledhill & Liu, 2015). In 2015, *TESOL International Journal* devoted an entire issue to SFL. In that special issue, Cheung (2015) and Humphrey (2015) analyzed voice and stance in ESL writing from the SFL framework; however, their papers were not related to the teaching aspect. The lack of practice and attempts of SFL-based pedagogy in L2 writing still remains a problem.

This paper reports on a teaching trial using SFL-based genre pedagogy (Rose, 2011), focusing on reporting verbs from the Interpersonal aspect of language use, to scaffold ESL learners in writing. Targeting problems in reporting verbs by a group of ESL student writers in a first-year writing program in a North American university, the author/instructor utilized a simplified system of Appraisal (Martin & White, 2005) and designed a two-week teaching intervention on reporting verbs, guided by the genre approach of modelling, joint construction and independent construction (Derewianka, 1999; Gibbons, 2002; Hyland, 2007; Rose, 2011). To test the effectiveness of this teaching trial, these students' original and revised short in-class writings and argumentative research papers were compared to see if they were able to deploy reporting verbs more effectively. These students were also anonymously surveyed at the end of the intervention for their perspectives of this pedagogical intervention and understanding of reporting verbs. The results indicated their improvement in deploying reporting verbs with some lingering problems, their enhanced understanding of reporting verbs and positive reactions to the teaching activities. Pedagogical implications for SFL-based approach in the EAP (English for Academic Purpose) context for ESL



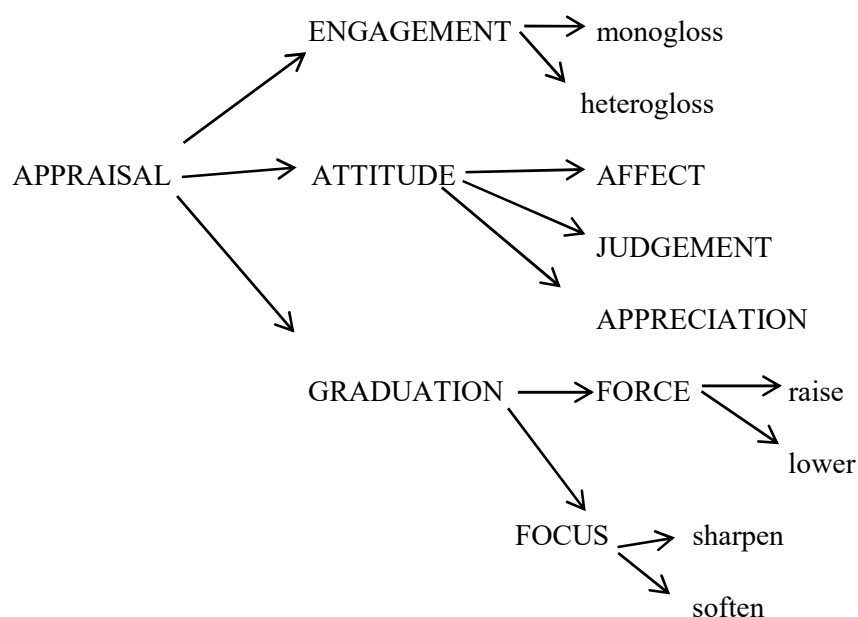
writing are discussed later.

### Literature Review

Halliday (1978, 1994, 2007) reiterated the importance of role relationship among language users under the Interpersonal Metafunction in SFL which conceptualizes language as meaning exchange activities. It focuses on how language constructs or negotiates relationship between users and expresses value judgments. Martin and White's Appraisal Theory (2005) furthers this line of study and is "concerned with the ways in which [language] resources... position the speaker/writer with respect to the value position being advanced and with respect to potential responses to their value position" (ibid:36). Here, we can investigate "the play of voices around opinions in discourse" (ibid:35) to see how attitudes are expressed. The Appraisal Theory comprises Attitude, which is on feelings, judgments and evaluations; Engagement, which is about sourcing attitude and voices; and Graduation, which is on the gradability of evaluation (pp. 35-37). Figure 1 presents a basic sketch of Appraisal system, based on Martin and White's 2005 figure.

Figure 1

*Sketch of the APPRAISAL Theory (adapted from Martin & White, 2005: 38)*



Considering complexity of this system and time allowed for teaching, the author/instructor simplified the Appraisal theory significantly for the targeted ESL student writers. Engagement was excluded in teaching; Attitude was simplified into "positive, neutral, negative" attitude towards what others say; and Graduation was simplified into three degrees of "strong, medium and weak". The idea of Attitude and Graduation was applied to reporting verbs to see how they could express author stance, writer attitude and interpretation towards other claims (Thompson & Ye, 1991). Following the common practice, "author" refers to authors of cited sources and "writer" refers to the writer of

the specific essay. Examples of reporting verbs showing positive attitude are “agree, support, concur, argue for, propose, confirm”, etc., and for negative attitude are “challenge, disapprove, disagree,

doubt, question, refute, dispute”, etc., and for neutral attitude are “study, examine, report, find, indicate, reveal, address, present, view, analyze”, etc. For Graduation, examples of strong reporting verbs are “stress, emphasize, affirm, insist, contradict, challenge, refute”, etc., and examples of weak reporting verbs are “suggest, imply, propose, hypothesize, speculate”, etc., and examples of medium reporting verbs are “show, express, list, comment, inform, discuss, mention”, etc.

Many student writers do not understand why sources are needed and they just use sources superficially to comply with teacher requirement, to earn high marks (Stockall & Cole, 2016) and to avoid plagiarism. Mori (2017:20) regretted that the current teaching on source integration was mostly “a lesson on plagiarism and the three forms of source incorporation”. Lee, Hitchcock and Casal (2018) echoed that surface forms of reporting verbs have been taught and studied more than stance. In reality, students need “to be taught how to create a ‘conversation among voices’” (Mori, 2017:20). When students fail to conceptualize writing as an interpersonal activity, they cannot appreciate the value of sources in building their own voices, which should be based on their critical thinking of other voices. Hence, they either do not use sources, or use sources in patchwork without making connections to their voices, or make claims as if those are their ideas when they are not (Wingate, 2012).

As one linguistic means for Appraisal, reporting verbs can facilitate creation and maintenance of connection between sources and enable writers to display their stance toward others (Hyland, 2007) such as mere presentation, support, criticism or question. Student writers could benefit greatly from this important tool to create interactions between themselves and sources they use in their own writing. Bloch (2010) created a corpus to teach reporting verbs, which included stance of reporting verbs as one important category. Though this is very useful as teaching materials, there is not much pedagogical support and suggestions in this area. Most studies on reporting verbs targeted at analysis of student writings, not on teaching (Charles, 2006a, 2006b; Kwon, Staple & Partridge, 2018; Lee, Hitchcock & Casal, 2018; Thompson & Ye, 1991). Using genre approach to teach reporting verbs with theoretical support from Appraisal Theory could be an effective way to help ESL student writers understand its value so that they can deploy reporting verbs more effectively.

Genre approach of the Sydney School chooses functional grammar as its linguistic framework, with an interventionist social goal to “redistribut[e] semiotic resources through education” (Rose, 2011:209). Its perspective “is social rather than cognitive”, and it is a “stratified, metafunctional, multimodal theory of text in social context” (ibid). Genre pedagogy features a teaching-learning cycle of “modelling, joint construction, and independent construction” (Derewianka, 1999; Gibbons, 2002; Hyland, 2007; Rose, 2011) to provide step-by-step scaffolding to students. In teaching writing, “modelling” is to “introduce a model of the genre to the class” to let them “become familiar with its purpose and features” by choosing or composing “a text which is similar to the one to be written later” (Derewianka, 1999:7). After that, students could “participate in the group writing of a text in the chosen genre” (ibid:8) in various forms in the stage of “joint construction”. Once students

understand the purposes, functions and features of the genre, the class can move on to “independent construction”. Gibbons (2002:67) added that the curriculum cycle was not limited to a single class and may take weeks, and the joint construction stage should be “teacher-guided” but not “teacher-dominated”. She provided suggestions such as introducing some meta-language in the modelling stage, eliciting student contribution in the joint writing stage, designing drafting/editing/peer editing/teacher conference activities in the independent construction stage in writing classes. Gibbons also effectively justified the need for explicit teaching and argued that “while imagination and ownership are important concepts in teaching writing, they are insufficient to ensure that all students, especially those less familiar with the language of school, will learn to write in a broad range of contexts” (ibid:68).

The current teaching trial on reporting verbs followed those suggestions by providing meta-language and explicitly teaching the simplified concept of Appraisal related to reporting verbs in the modelling stage, eliciting students’ input in the joint construction stage and designing multi-step assistances in the independent construction stage. The ultimate purpose of this teaching trial is to test if ESL student writers could deploy reporting verbs more effectively when they receive instruction on how to use language strategically to create an interactive sense of writing.

## **The Pedagogical Trial**

### **Background**

Being a college composition instructor for ESL students for many years, the author has repetitively witnessed students’ frustration at source incorporation. One common problem is the students’ monotonous choice of “say, write, think” as the main reporting verbs when quoting others’ ideas. As many student writers are unable to use reporting verbs meaningfully to create the interactional and dialogic sense among sources, I decided to implement a pedagogical intervention on this. SFL provides powerful linguistic tools to explain the interactional nature of language use in writing, and the SFL-related genre pedagogy has been proven to be effective in language teaching (Hyland, 2007; Rose, 2011). So they served as the theoretical frame and teaching method for the designed activities. Due to time concern, a simplified version of Appraisal Theory (Attitude and Graduation) was adapted to teach reporting verbs.

This teaching trial was conducted in a first-year ESL composition class in a North American University. In the composition program, ESL students need to complete the same writing tasks as L1 English writers in other classes, one of which is the research argumentative paper (hereafter RA), in which they must express their stance on a controversial issue, using support from outside sources. As the instructor, I value process writing and require students to take several steps from generating ideas, outlining, drafting, getting feedback to revising before they turn in the final papers. In the semester of this pedagogical trial, there were 16 ESL students from four continents with various L1s, ranging from early 20s to 30s in age. They have all met the university admission requirement in English proficiency, and have taken the first composition course in the previous semester at the same

university. This 2-week/4-class pedagogical intervention adopted the curriculum cycle of modelling, joint construction and independent construction from the genre approach (Derewianka, 1999; Gibbons, 2002; Hyland, 2007; Rose, 2011).

### Sequence of Activities in this Teaching Trial

In order to test the effectiveness of this teaching trial, other than the teaching activities, a draft and revised version of a short in-class writing and a long researched argumentative essay from each student were collected. The purpose was to examine students' deployment of reporting verbs before and after instruction as evidence for the (in)efficiency of this teaching trial. Table 1 presents the sequence of all activities in this teaching trial.

Table 1

*Sequence of activities in the teaching trial*

Sequence	Materials Collected	Teaching Activities
before the teaching trial	draft 1 of RA essay	no
class 1	draft 1 of short in-class writing	students write in class to express their views on a topic
class 2	No	explicit instruction on reporting verbs from the Interpersonal perspective
class 3	No	modelling and joint construction of deploying reporting verbs at individual sentence level
class 4	No	modelling and joint construction of deploying reporting verbs at paragraph level
after the teaching trial	survey on teaching and reporting verbs revised short in-class writing and RA essay	no

### Collecting and Analyzing Students' Writing Samples and Survey

To ensure validity and reliability in analyzing students' writing samples, each student was assigned a number, and they were told to put the same number, instead of their names, on their writings. All sentences in their writings containing sources and/or reporting verbs were hand-picked for analysis. The author kept two clean sets of copies, and separately coded reporting verbs in each paper twice, with a 3-week interval in between. The categories of Attitudes (Positive, Negative and Neutral) and

Graduation (Strong, Weak and Medium) were found to be the same in both codings, and the same reporting verbs were identified both times. This ensured the data coding reliability. An anonymous survey was issued at the end of the teaching intervention on students' perceptions of the teaching activities and their understanding of reporting verbs. The survey had several open-ended questions on students' current understanding of reporting verbs, and their responses to what worked and what did not work in the teaching activities. Two sample sentences with the same quotation but different reporting verbs were also included in the survey to get students' evaluations on different reporting verbs to further probe their understanding of the value of reporting verbs.

## **Teaching Activities**

### **Before the Intervention**

Before the teaching intervention, these students had already received instruction on Modern Language Association citation for works-cited page and in-text citation format. When asked why they need to cite other people's ideas, most responded that it was required to avoid plagiarism. A few mentioned that this could support their own points. Most ESL students in this class had little idea of why they should document sources. Their researched argumentative essay draft 1 were collected and their deployment of reporting verbs were analyzed. The results will be later presented in Table 3.

## **Pedagogical Intervention Sequence**

### **Class one**

In class one, students were asked to complete an in-class short paragraph writing on a controversial issue. Three short readings with different stances on the issue (positive, negative, neutral) were provided to the students, and students were asked to use all of them to help them argue for or against the issue in one or two paragraphs. The instructor did not explain the content and stance of those reading materials. This short assignment was designed because students in Doolan and Fitzsimmons-Doolan's (2016) study could not perform long assignments as successfully as minor ones in source incorporation. I was interested to see if my students would demonstrate the same behavioral differences in short writings and long writings for reporting verbs, and if so, the pedagogical implications of this. Analysis of reporting verbs in their short in-class writings will be presented in Table 2 later.

### **Class Two**

Class two was explicit instruction on source incorporation and reporting verbs. The class started with a hypothetical conversational scenario of the students wanting to join a debate. Two strategies were offered: one was to cut in directly with their opinion; the other was "I have been listening to you for some time, and I understand what is being discussed. Here is what I think." Unanimously, the students chose the second one as it gave them legitimacy to participate in the conversation. This idea was then extended to writing to help students understand that writing relies on the same concept of

interpersonal interaction between writers, readers and other authors. The instruction emphasized that to write effectively, writers can benefit from deploying a variety of reporting verbs to represent other sources, respond to other sources with critical evaluation and make sources talk to each other. Students were then shown several citations led by different reporting verbs with the same content, and they were asked to judge if those different reporting verbs influenced their interpretation of the author's attitude towards what was cited: "Smith says/announces/challenges/supports/explains that...". Students were able to gauge the effects of different reporting verbs successfully. The explicit teaching of reporting verbs introduced metalanguage of Stance (negative, neutral, positive) and Graduation (strong, medium, weak) from the Appraisal Theory, hoping to make the idea easily accepted within the class hour by those students. In the explicit teaching, students were also provided with a list of common reporting verbs and were directed to categorize those verbs into positive, neutral, negative groups, first individually then in groups. Then the whole class checked the answer together. The same activity was used for graduation of reporting verbs. This explicit instruction aimed at enhancing student's knowledge of usage and function of reporting verbs to appreciate how reporting verbs could strategically reflect various stances in writing.

### **Class Three**

Class three modelled using reporting verbs for sources at the sentence level, followed by joint construction. The class started with an in-depth analysis of one of the three previously provided readings. The whole class examined some selected words/sentences, which reflected the stance of the writing and functions of the selected parts- illustrating, defining, discussing, showing results, providing details, expressing ideas, explaining, showing different views, etc. Then different reporting verbs were explained to show how to represent others' ideas at the sentence level in a more meaningful way. After three modelling examples, students were encouraged to provide reporting verbs on their own for other sentences. When we finished practicing different reporting verbs for this reading, the class moved on to the joint construction stage and the students followed suit to represent ideas in the other two readings by choosing different reporting verbs at the sentence level in groups. Finally, each group presented one sentence to the class for peer feedback on effectiveness of their reporting verbs and sentences.

### **Class Four**

Class four focused on using reporting verbs at the paragraph level to make use of multiple sources, followed by joint construction to teach students how to integrate various sources in writing. To help them see the interaction of sources, students first compared and linked ideas in the three readings: similar and different ideas/sentences, sentences as evidence/example, sentences that express author point, sentences that explain, etc. Afterwards, the class examined a written paragraph on this topic, which used these three readings, to analyze if reporting verbs were used properly across sentences to reflect source stance and create interactions between sources in the entire paragraph. Good usages

were selected together with bad usages of reporting verbs. Students then paired up to revise ineffective use of reporting verbs and sources in that sample paragraph. The final activity discussed effectiveness of students' revisions on reporting verbs and their reasons for revisions. Class four was the last class in this pedagogical intervention.

### **After the intervention**

After class four, students were asked to revise their previous in-class short writings and researched argumentative essay, focusing on source incorporation and reporting verbs, which were later collected to examine usage of reporting verbs. There was no extra help for them to revise the short in-class writing, but they went through peer review and individual conference with the instructor to revise their RA essays. According to the provided peer review guideline, students had to evaluate source incorporation and reporting verbs in peer writing. The instructor also discussed at least one instance of their source usage and reporting verbs with the students in the individual conference session. Other than the two revised writing tasks, students were also issued a survey to get their perceptions of the effectiveness of the teaching activities and their current understanding of reporting verbs.

### **Analysis of reporting verbs in writings before and after the pedagogical trial**

Before the teaching trial, students' RA paper and short in-class writing were collected. After the teaching trial, their revised short in-class writings and revised RA paper were collected to analyze their use of reporting verbs. The small data size does not suffice for statistical analysis, so only descriptive quantitative data of reporting verbs in these four sets of writings are reported. Information of total number and percentage of reporting verbs and the Attitude and Graduation of reporting verbs are presented. Reporting verbs that were used once or twice in the data are defined as low frequency, those of three or four times are defined as medium frequency, and those of five times and more are defined as high frequency. Percentage is calculated as the number of each category divided by the total number of word count. Table 2 compares the descriptive data of reporting verbs in the draft and revised version of the short in-class writing. Table 3 compares the descriptive data of reporting verbs in the draft and revised version of the researched argumentative paper.

Table 2

*Descriptive Data of Reporting Verbs in In-class Writings: Draft and Revised*

<b>Draft</b>			<b>Revised</b>		
word count #	reporting verbs #	% of reporting verbs	word count #	reporting verbs #	% of reporting verbs
3884	38	0.97%	4775	58	1.2%
high frequency	medium frequency	low frequency	high frequency	medium frequency	low frequency
22	6	10	13	21	24
0.56%	0.16%	0.25%	0.27%	0.44%	0.5%
<b>Stance</b>			<b>Stance</b>		
Positive	Negative	Neutral	Positive	Negative	Neutral
1	0	37	13	3	42
0.03%	0%	0.95%	0.27%	0.06%	0.88%
Graduation			Graduation		
Strong	Medium	Weak	Strong	Medium	Weak
2	34	2	14	39	5
0.05%	0.87%	0.05%	0.29%	0.82%	0.1%

Quantitatively, in the draft of the short in-class writing, although students used medium-frequency reporting verbs “state and explain” (3 each) and some low-frequency reporting verbs such as “suggest, see, point, indicate, report, compare, prove, believe, imply” (10 in total), they highly relied on two reporting verbs to introduce sources: “say” (12) and “show” (10). This reflected their limited choice and lack of awareness of the value of reporting verbs. Their reporting verbs dominantly expressed a neutral attitude and a medium degree of that attitude (37 and 34 out of the total 38 reporting verbs respectively). These clearly distanced student writers from the sources without showing their stance and commitment, which did not help them engage in meaningful conversations with other sources. Overall, this reflected their low competence in this area. In comparison, in their revised short in-class writings, other than two high frequency reporting verbs of “show” (8) and “emphasize” (5), they used more diverse reporting verbs such as “confirm, support, point” (4 each), “say, argue, claim” (3 each), “explain, suggest, argue against, state” (2 each), and “prompt, imply, report, defend, compare, affirm, endorse, echo, indicate, review, prove, declare, illustrate, reason, discuss, cast doubt on” (1 each). Take the word “say” for example, in the draft, it was used 12 times in total and that number dropped to 3 in the revised version, which indicated students’ efforts to use other more effective reporting verbs to replace it. Though students’ reporting verbs still largely showed their neutral stance (42 out of 58 reporting verbs), they were also able to show positive stance (13) and negative attitude (3) more often, as those numbers apparently increased from the 1 instance



of positive attitude and zero instance of negative attitude in the draft. Similarly, even though they still relied on medium degree reporting verbs (39 out of 58 reporting verbs), the instance of strong attitude increased from 2 in the draft to 14, and that of weak attitude increased from 2 to 5 in the revision.

Qualitatively speaking, their reporting verbs in the revised version indicated both positive stance “confirm, support, claim, affirm, endorse, prove” and negative stance “argue against, cast doubt on”. Students also differentiated strong reporting verbs “emphasize, declare, confirm” from weak reporting verbs “suggest, imply” in the revision. Some students indicated functions of the cited ideas with reporting verbs such as “compare, review, illustrate, reason, explain”, and interactions between different sources, such as “defend, endorse, echo, confirm, argue against, cast doubt on”. Many students created dialogues between sources in their revisions. Some examples are provided here, all taken from the students’ revised version:

1. While Schlichter prompts the need..., Leung implies that... (student 3)
2. Sager emphasized the need for.... This is then echoed by the FBI statistics ... (student 6)
3. Different from Sager’s ideas, Schlichter argues against this... (student 7)
4. Sager argues that..., but Leung defends that... (student 11)
5. This is one of the supporting reasons.... Another supporting statement is endorsed by Schlichter ... (student 12).
6. We see that Leung emphasizes that..., but I do not agree with this statement because... (student 14).

All these evidence these students’ progress in deploying reporting verbs to properly indicate attitude of sources to create dialogues in their revised writings. It is indeed inspiring to witness such obvious improvement in students’ writings just after a short period of instruction.

Will students improve usage of reporting verbs in a longer and more demanding writing assignment when they need to pay more attention to bigger issues of content and structure? Will the teaching trial help students to actually employ this knowledge in writing? To answer that question, students’ researched argumentative drafts (collected before the teaching trial) and their revised RA paper (after the teaching trial) were compared for their use of reporting verbs. Table 3 presents this comparison in the draft and revised version of their RA paper.

Table 3

*Descriptive Data of Reporting Verbs in Researched Argument Paper: Draft and Revised*

<b>Draft</b>			<b>Revised</b>		
word count #	reporting verbs #	% of reporting verbs	word count #	reporting verbs #	% of reporting verbs
24653	69	0.28%	38583	156	0.4%
high frequency	medium frequency	low frequency	high frequency	medium frequency	low frequency
56	0	13	97	31	28
0.23%	0%	0.05%	0.25%	0.08%	0.07%
Stance			Stance		
Positive	Negative	Neutral	Positive	Negative	Neutral
1	0	68	21	1	134
0%	0%	0.28%	0.05%	0%	0.35%
Graduation			Graduation		
Strong	Medium	Weak	Strong	Medium	Weak
1	68	0	26	119	11
0%	0.28%	0%	0.07%	0.31%	0.02%

Quantitatively speaking, these students used fewer reporting verbs in their RA draft, and relied mostly on a few high frequency neutral attitude, medium degree reporting verbs: “say” (24), “state” (16) and “show” (16). Other reporting verbs of “find, argue, conclude, point” (2 each) and “note, report, add, mention, believe” (1 each) were used at low frequency. In the draft, out of the 69 reporting verbs, “believe” was the only one showing positive stance and strong attitude; the rest were all neutral, medium degree reporting verbs. It is clear that students’ use of reporting verbs in the draft was limited and ineffective. In comparison, their reporting verbs increased from 69 to 156 in the revision, and were more diversified: “say” (17), “state, show” (15), “confirm” (11), “add” (10), “claim” (9), “argue, explain, report, estimate” (5 each), “note, discuss, observe, find” (4 each), “highlight, mention, conclude, emphasize, imply” (3 each), “see, suggest, assert, support, define” (2 each), and “announce, prove, believe, compare, command, raise an interesting point, reveal, worry, survey, point out, stress, put, quote, agree, comment, defend, insist, declare” (1 each). More reporting verbs in the revision showed positive stance “confirm, assert, support, prove, believe, raise an interesting point, insist, declare, agree”, which was an apparent increase from the one positive stance in the draft. In the draft, there was no negative stance reporting verb, and in the revision, there was only one such verb “worry”. Possible explanation for this might be that students focused more on finding sources that support rather than contradict them, hoping to make their argument stronger that way. They also used more reporting verbs showing strong (26) and weak degree (11) in the revision; in the draft, there was only one reporting verb showing strong degree. So even though most of their reporting verbs still showed neutral stance and medium degree, the students indeed deployed more diversified reporting verbs and

showed various stances and degrees.

Qualitatively speaking, “say, show, state” still remained as their top choice of reporting verbs, which indicated some lingering issues of reporting verbs in a more demanding writing. When their attention was redirected to larger concerns of argument, content, structure, they could not spare time or energy to carefully select specific and proper reporting verbs, unlike what they did in the short in-class writing when they were not cognitively over-taxed. Despite that lingering problem, students have chosen a much wider variety of reporting verbs to indicate positive and negative stances and strong and weak degrees more often. Some of their reporting verbs were not even in the provided list to them. Students could also indicate functions of cited ideas and create dialogues with reporting verbs. Some examples are presented here, all taken from the revised RA essays:

7. According to Adam’s article, people think... This statement agrees with the controversial issue. (student 3)
8. Reece asserts that.... His paper is not an anomaly: there are many studies linking ... (student 6)
9. Kataguiriri ... argued that .... Besides that, according to the Institute of Geography and Statistics, the latest population research confirmed that... (student 7)
10. Wanyama agrees with Karachi and Arowolo that ..., as he comments that... (student 12)
11. From his article I learnt that ... when he states that ... (student 4)

Overall, students have demonstrated improvement in reporting verbs and source incorporation in both revisions of the short in-class writings and RA essays, and the changes of reporting verbs were more noticeable in the revised short writings.

### **Analysis of Survey data**

After they submitted their revised RA essays, those sixteen students were surveyed anonymously to probe their current understanding of source incorporation, reporting verbs and the usefulness of the pedagogical intervention. Compared with their initial responses of why sources are needed in writing, more students mentioned the following reasons: supporting ideas (12), giving credible argument (12), adding reliable information/evidence (11), seeing differences between ideas (10), gaining credibility as a writer (9), backing up own writing (9), avoiding plagiarism (9), directing interested readers for more readings (6), making the paper stronger (6), and showing other ideas (5). When asked about how they chose ideas to be cited in their writing, ten students answered that they would read short articles more carefully but glance over long articles to locate sentences for possible quotations, and five students indicated they spent a lot of time reading and comparing different sources. For the importance of understanding and indicating attitudes of outside sources, 14 agreed that it was important as they need to mention (counter)argument in writing, to see things from different views, to understand why others disapprove or support an idea, and not to misinterpret others’ works. Only

two students felt it was not important to do so as their ideas did not depend on other authors' attitudes. 13 students emphasized that they relied on reporting verbs to indicate source stance and even listed some reporting verbs as examples in their answers.

The students were also given two sentences to probe their preference for reporting verbs. Sentence A was "X mentions that 'a direct quotation'". B was "There are many ways to teach pronunciation. One of such methods, as suggested by X, is 'the same direct quotation'". 11 students chose B because "it has an introduction of the idea and prove it with a quote", "it connects the writer's idea to the outside source's idea", and "it is the correct way to use outside source". One student explained that "B sounds convincing and the reader will understand the writer stance. In A, we do not know if the writer agrees or disagrees". These showed their understanding of interaction between sources and writer ideas. Five students chose A because it was easy to read, which was somewhat unexpected and indicated different student understandings. The final question was on their perception of the pedagogical intervention. All but one student answered that it was useful/helpful, as it opened their eyes to different attitudes they did not know before; it helped them understand the role of reporting verbs so that they could use various reporting verbs rather than repeat the same reporting verbs; and the revision helped them gain confidence when they saw how much better they could perform. Seven students wished for more similar activities and more detailed explanations. One student did not feel it helpful as he/she "expect[ed] the professor to go around in the class working with each and every student, so they use the sources correct and not just telling the students in a big class". This questionnaire clearly indicates students' enhanced awareness and better knowledge of source incorporation and reporting verbs, proving the effect of such a short-term pedagogical intervention on reporting verbs. Most of the students have benefited from the teaching intervention, which has broadened their mind to the interactive nature of writing and the important roles of reporting verbs.

## Discussion and Conclusion

In order to become a better communicator, ESL student writers need to conceptualize writing as an interactive activity. They should develop the knowledge that stance representation is quintessential in writing, and reporting verbs are instrumental to do so to create interactions. Those students' short writings and RA drafts before the pedagogical intervention reflected their lack of knowledge of source incorporation and how reporting verbs could help them write a better paper to communicate with other sources and their readers. Their revised short in-class writings and RA essays strongly evidenced the effectiveness of explicit instruction on reporting verbs, as their writing samples have demonstrated noticeable improvement in source incorporation and reporting verbs. The final survey data also confirmed the effectiveness of this teaching trial which adopted the genre approach. Genre approach has been shown to be effective in teaching students the purposes and stages in different writings (Hyland, 2007; Rose, 2011); this pedagogical trial indicated that it was also helpful to focus on some particular aspects of writing. Since helping students understand the functions of citations is

much needed in composition classes (Lee, Hitchcock & Casal, 2018), the genre approach could be adopted in ESL writing classes to achieve that purpose.

Mori (2017:7) represented a group of scholars and classroom practitioners' belief in the power of helping students view language as fundamentally dialogic and understand "how speakers evaluate ideas, express stance towards individual, and overall interact with the world and the people". Mori's suggested activities of discussion of stances, categorization of verbs, experiment with verbs to see their appropriateness and comparison of student drafts were included in this reported pedagogical trial. When students could conceptualize writing as dialogues between themselves, sources and readers, they would value proper deployment of reporting verbs and sources as an interpersonal strategy to help them build their voices. First-year student writers in composition classes learn writing skills to prepare them for future content-based academic writing, which definitely requires expressing and building their voices on sources. If our students leave our classes with a superficial understanding of the surface format of source usage without understanding the essential reasons of why and how sources could help them express their own voices in their future communities, we are not doing our job to help them face those challenges. Halliday's Interpersonal metafunction makes explicit understanding of this aspect of language use easier, and the Appraisal theory provides explanatory tools to teach students how to deploy linguistic means to create better interpersonal interactions in writing. Hopefully, this brief pedagogical trial adds evidence to the effectiveness of how SFL-based teaching could enlighten students who are not familiar with the genre and convention of writing to understand the function and reasons behind their writing choices in the academic context.

The pedagogical cycle of modelling, joint construction and independent construction in the genre approach (Derewianka, 1999; Gibbons, 2002; Rose, 2011) also provides effective scaffolding to student writers. One quintessential goal of genre approach is to redistribute semiotic sources through education (Rose, 2011) and help disadvantaged students master those sources. Writing in the EAP context is one such context where the genre and writing convention is unknown to outsiders and challenging to novices, which desperately needs to be decoded to new members. Not every novice writer can succeed on their own; and even when they do, this is a long, hard process through failures. Applying genre approach in teaching writing for the academic purposes could be a shortcut to the students, and the teaching effectiveness of genre approach in the language teaching field outside the EAP context has already been proved (Hyland, 2007; Rose, 2011). In this teaching trial, the modelling stage provided explicit instruction in conjunction with simplified idea of the Appraisal Theory to help student writers understand the concept. The collaborated stage elicited students' input to work with each other and with the instructor to improve the writing in terms of reporting verbs. These two steps have ensured students' improved performance in reporting verbs in their individual writings later. Analysis of students' original and revised writings (long and short) in reporting verbs before and after this teaching trail added evidence to the effectiveness of such a pedagogical approach in the EAP context.

Like Doolan and Fitzsimmons-Doolan's (2016) participants who performed better in minor assignments than major assignments after an 8-week teaching, my students have also made more apparent progress in the short writing assignments in source incorporation and reporting verbs than the long, more demanding argumentative paper. They still showed some lingering problems of reporting verbs in the longer argumentative essays. Several reasons could explain this somewhat disappointing behavior: students have spent much more time discussing the sources for the short writing than the longer argumentative essays; students have received extensive help in the short writings but not as much in writing the argumentative essays; there was much less to consider in the short writing than in the long argumentative essays (overall structure, coherence, content, more sources); writing skills accumulate and improve gradually over time, this one-time two-week pedagogical trial is not enough for students to master the skills completely. So in order to better facilitate ESL student writers to create proper interaction and presentation of sources in their writings, this knowledge needs to be revisited frequently, and students need more help along the way to finally master this skill. This is another important pedagogical implication from this teaching trial. In line with this, regretfully, I was not able to collect those students' writings in the following semester to examine the longitudinal effect of this teaching trial on students' deployment of reporting verbs.

McCabe, Gledhill & Liu (2015) lamented at the lack of SFL-related pedagogy in language teaching and called for more studies on that. The teaching trial reported in this paper answered that call and shedded light on how genre approach, in conjunction with SFL knowledge, could help student writers to a great extent in the EAP context of writing. This pedagogical trial only lasted two weeks. Given more time, repeated instruction on the same skill or instruction on other aspects of source incorporation could be explicitly delivered in class to better help ESL students appreciate that writing is not a monologue but a carefully-crafted dialogue that enables them to interact with various related parties, to talk to others more convincingly, and to make themselves more credible writers.

## References

- Bloch, J. (2010). A concordance-based study of the use of reporting verbs as rhetorical devices in academic papers. *Journal of Writing Research*, 2(2), 219-244.
- Charles, M. (2006a). The construction of stance in reporting clauses: A cross-disciplinary study of theses. *Applied linguistics*, 27 (3), 492-518.
- Charles, M. (2006b). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes*, 25, 310-331.
- Cheung, L. M. E. (2015). Legitimizing the knower's multiple voices in applied linguistics postgraduate written discourse. *TESOL International Journal*, 10 (1), 62-76.
- Coffin, C. (2009). Incorporating and evaluating voices in a film studies thesis. *Writing & Pedagogy*, 1, 163-193.
- Derewianka, B. (1999). *Exploring how texts work*. Sydney: Primary English Teaching Association.
- Doolan, S. M., & Fitzsimmons-Doolan, S. (2016). Facilitating L2 writers' interpretation of source texts. *TESOL Journal*, 7 (3), 716-745.
- Gibbons, P. (2002). *Scaffolding language scaffolding learning: Teaching second language learners in the mainstream classroom*. NH: Heinemann.
- Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. London: Arnold.
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London: Arnold.
- Halliday, M. A. K. (2007). *Language and Society. Volume 10 in the collected works of M.A.K. Halliday*. J. J. Webster (Ed.). Peking: Peking University Press.
- Halliday, M. A. K., & Hasan, R. (1985). *Language, Context, and Text: Aspects of language in a social-semiotic perspective*. Oxford: Oxford University Press.
- Humphrey, S. (2015). Building a critical stance in academic and civic discourse: Burnishing and tarnishing. *TESOL International Journal*, 10 (1), 47-61.
- Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, 16, 148-164.
- Kwon, M. H., Staples, S., & Partridge, R. S. (2018). Source work in the first-year L2 writing classroom: Undergraduate L2 writers' use of reporting verbs. *Journal of English for Academic Purposes*, 34, 86-96.
- Lee, J. J., Hitchcock, C., & Casal, J. E. (2018). Citation practices of L2 university students in first-year writing: Form, function, and stance. *Journal of English for Academic Purposes*, 33, 1-11.
- McCabe, A., Gledhill, C., & Liu, X. (2015). Systemic functional linguistics and English language teaching. *TESOL International Journal*, 10 (1), 1-10.
- Martin, J. R. & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. New York: Palgrave Macmillan.

- Mori, M. (2017). Using the appraisal framework to analyze source use in essays: A case study of engagement and dialogism in two undergraduate students' writing. *Functional Linguistics*, 4 (11), 1-22.
- Rose, D. (2011). Genre in the Sydney School. In J Gee & M Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (209-225). London: Routledge.
- Stockall, N., & Cole, C. V. (2016). Hidden voices: L2 students' compensatory writing strategies. *Teaching in higher education*, 21 (3), 344-357.
- Thompson, G., & Ye, Y. (1991). Evaluation in the reporting verbs used in academic papers. *Applied Linguistics*, 12 (4), 365-382.
- Wingate, U. (2012). 'Argument!' helping students understand what essay writing is about. *Journal of English for Academic Purposes*, 11, 145-154.



## Taiwan University Students' Perceptions of Summative and Formative Classroom Assessment in English Courses

Yi-Ching Pan\*

*National Pingtung University*

### Abstract

Classroom assessment is an essential component of teaching and learning. Still, the literature on language testing often highlights teachers' perceptions of designing classroom assessments with little concern about students' perceived realities. This study thus explores Taiwan university students' perceptions of summative and formative classroom assessment tasks in their English courses. To address this issue, 107 first-year undergraduates at one Taiwan university were recruited to complete one summative written exam and two formative assessments, after which they filled in a questionnaire consisting of six subscales: congruence with planned learning, authenticity, student consultation, transparency, diversity, and washback effects. We discover four major findings: 1) students were in favor of the summative assessment due to its congruence with planned learning and transparency; most students admitted that they learn more from preparing for the summative assessment; 2) students viewed the cooperative group assessment positively, because of its diversity; 3) preparation for summative assessments elicited a greater degree of test-oriented learning for respective skills, whereas formative assessments enhanced students' motivation to learn English for productive skills; and 4) students believed that an appropriate combination of summative and formative assessment tasks benefits their learning. The findings provide further pedagogical implications.

**Key words:** consequences of test use, washback, classroom assessment, summative assessment, formative assessment

---

\* [yichingpan@yahoo.com.tw](mailto:yichingpan@yahoo.com.tw)

## Introduction

Since the literature has increasingly scrutinized the consequences of tests, many studies have investigated the washback on large-scale standardized language tests to determine the effects brought about by educational policies in different countries. For example, studies by Hayes & Read (2003), Ferman (2004), Gebril & Brown (2014), Cheng (2005), Qi (2005, 2007), Xie (2015), Zhan & Andrews (2014), and Smyth & Banks (2012) explored how high-stakes entrance examinations affect teaching and learning in New Zealand, Israel, Egypt, Hong Kong, China, and Ireland, respectively. Wall & Horak (2006, 2008, 2011) and Green (2007) analyzed the impacts of TOEFL (Test of English as a Foreign Language) and IELTS (The International English Language Testing System) on teaching in European ESL (English as a second language) classes Shih (2007), Tsai & Tsou (2009), Gan (2009), Pan, (2014, 2018), and Xie (2013) investigated the positive and negative effects brought about by English exit requirements using standardized language tests in Hong Kong and Taiwan higher-education institutes. These aforementioned traditional large-scale standardized tests, also known as assessment of learning (Earl, 2003), usually focus on learning outcomes, yet hardly address learning processes. Different types of assessments, such as assessment for learning and assessment as learning (Earl, 2005), have therefore been subjected to heightened levels of consideration. In addition to the raised awareness of multiple methods of assessments, many countries have proposed the utilization of standardized tests due to their fairness and reliability, following pressure to seek educational accountability and quality control (Black, 2000). The current trend is to adopt different types of classroom assessments as a fair and reliable means of monitoring students' learning processes and evaluating their classroom performance.

Assessments play an essential role in teaching and learning because they are crucial for informing teachers about their work, while at the same time, if the assessment is improved, then the resultant learning can be enhanced. A few studies exist that focus on the washback or impact of high-stakes testing of teaching and learning, especially from teachers' perceptions (Cheng, 2008, 2014; Spratt, 2005; Wall, 2000). However, very little research has explored how classroom assessments influence learning from students' viewpoints. To bridge this gap in the literature, we investigate Taiwan university students' perceptions of summative and formative classroom assessment tasks in their English classes and how these classroom assessments affect their learning.

## Context of the study

Students in Taiwan are required to learn English from the time they are in elementary school so that by the time they attend a university, they will have learned English for at least eight years. Many students are not very motivated to learn English, probably because Taiwan's education system is test-oriented and teacher-centered. At the university level, all first-year students are required to take two-three hours of English classes every week. Moreover, the majority of these students have stated that they are taking English classes, because they want to earn credits for graduation, with around 30% of these students passing the CEFR A2 level and 50% of them not passing it. In other words, quite a few

students had insufficient English proficiency, even though they had learned English for more than eight years.

### **Literature review**

This section reviews (1) the historical perspectives of washback, focusing on the contributing factors for various degrees of test effects and (2) the possible dimensions for understanding test-takers' perceptions of classroom assessments. The review on washback effects and assessment tasks serves as a theoretical basis for the purpose of this study to explore both students' perceptions of assessment tasks and the influence of the assessments on learning.

#### **The historical perspectives of washback focusing on the contributing factors for various degrees of test effects**

Hughes' (1993, 2003) trichotomy of the backwash model describes test effects in terms of "participants", "process(es)", and "product(s)". This model explains how participants interpreting and reacting toward a test affect how and what their responses to it will be, indicating that the quality of a test is essential to predict the degree of washback test effects.

Alderson and Wall (1993) propose 15 washback hypotheses and illustrate various possible effects brought about by tests primarily on teaching and learning, ranging in detail from the most basic to quite specific effects, as listed in Appendix 1. Alderson and Wall's hypotheses, like Hughes' model, address what washback effects might look like (i.e. the consequences) more than they focus on what factors other than a test (i.e. the mechanisms) lead to these effects. Nevertheless, several of the hypotheses imply that, in addition to a test, there might be some other factors that elicit more effects from some learners and teachers than they did for others. Not only does the quality of a test affect teaching and learning, but how a test is used under different situations also affects the strength of the washback, as in Hypothesis 13: "Tests that do not have important consequences will have no washback" (ibid.).

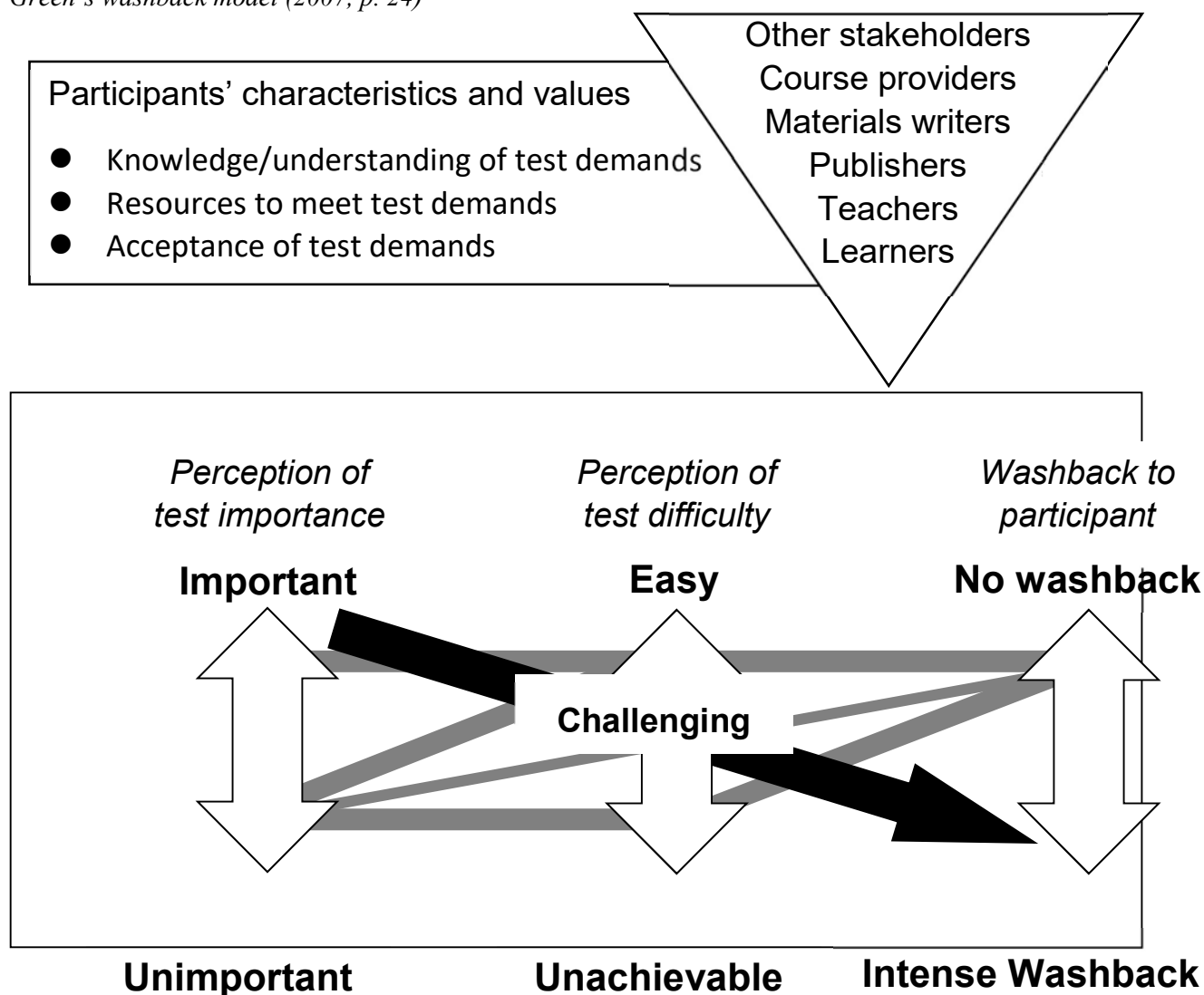
Green (20007) presents the concept of washback variability (see Figure 1), in which participants' perceptions of test stakes, test quality, and test difficulty tend to vary from person to person and therefore lead to differences in the washback effects experienced by individuals. In Green's opinion, students' perceptions, attitudes, and reactions toward the following seven factors may result in various degrees of washback intensity: (a) test demands (content, format, and complexity), (b) purpose of the test, (c) stakes of the test, (d) difficulty of the test, (e) test preparation resources, (f) teaching methods, and (g) learning content.

In review of the aforementioned studies, students' perceptions of classroom assessment can be explored from washback perspectives to understand how the assessment influences students in various ways. Using the above three washback models as the theoretical base for the questionnaire in

this study, Table 2 lists 11 items to investigate to what extent various assessment-related factors influence students' learning.

Figure 1

*Green's washback model (2007, p. 24)*



### The possible dimensions for understanding test-takers' perceptions of classroom assessments

Dorman and Knightly (2006) develop an instrument called Perceptions of Assessment Tasks Inventory (PATI) to observe students' perceptions of assessment tasks under five categories to understand the qualities of assessments, which served as the theoretical basis for developing the 12 items in the student questionnaire (see Table 2). These five categories are described below:

These five categories are congruence with planned learning, authenticity, student consultation in the assessment process, transparency about the purposes and forms of the assessment, and diversity.

Congruence with planned learning denotes that students perceive the assessment tasks as being aligned with their learning objectives and activities. Authenticity means that the assessment tasks are related to students' daily life. Student consultation implies that students are involved and consulted in the assessment process. Transparency signifies that students are well informed about the purposes and forms of the assessment. Diversity is when students perceive that they can finish the assessment tasks at their speed.

The overall assessment can be categorized into two parts. One is summative assessment whose goal is to *evaluate student learning* at the end of the semester by comparing it against some standard or benchmark. Examples include traditional paper-and-pencil tests such as mid-term and final exams and proficiency tests.

Summative assessment usually lacks feedback or any suggestions to improve performance (Brown, 2004). The other is formative assessment, whose goal is to *monitor student knowledge and understanding in order* to provide ongoing and useful feedback that can be used both by instructors to improve their teaching and by students to improve their learning. Much of the classroom assessments such as oral discussion, group/pair work, and completing a portfolio can be formative because students form their knowledge by analyzing and internalizing teachers' comments (Brown, 2004). However, it seems that formative assessment has not always been the focus of attention in ESL/EFL studies, especially in a test-oriented educational environment. If the students are not very proficient in English, as mentioned in the context of study, and they have become accustomed to summative written assessments, then a question arises: Would they hold more favorable opinions toward summative assessments than formative assessments, or the other way around? We shall investigate this to help fill the gap in the field of language testing.

## Research questions

Based on the literature review, this study thus explores two research questions.

1. What are the differences between students' perceptions of summative and formative assessments?
2. What are the differences between students' perceptions of summative and formative assessments based on their levels of proficiency?

## Method

### Subjects

One hundred and seven first-year students at one university were recruited for this study, of which 29 were male, and 78 were female. At the time of the study, they were taking a required 3-hour English class every week. Based on their mid-term exam scores, they were split into two groups: 54 students whose scores were in the top 50% were in the high-proficiency group, whereas the others (53) were in the low proficiency group.

## Instruments

The instruments utilized include three assessments, one questionnaire, students' learning reflections, and semi-structured interviews. This study reports mainly quantitative data due to length limitations.

### Three Classroom Assessments

The three assessments, including one summative and two formative assessments, were all designed

based on lecture material from the English reading class and were completed within a timeframe of one semester. The summative assessments consisted of two traditional multiple-choice listening and reading questions. As for the formative oral presentations, students were given a choice to select one question out of two and make a two-minute presentation. For the formative group audio PowerPoint presentation, students were asked to form groups of two to three people, create an audio PowerPoint, and write reflections based on what they had learned from this project. The students were also told in detail about the guidelines for the assessments, the purposes of completing these two formative assessments, when they should finish, and how they would be graded.

### Questionnaire

The questionnaire consists of two parts. The first part adopts Dorman and Knightley's (2006) Perception Inventory and includes five sections: Planned Learning, Authenticity, Student Consultation, Transparency, and Diversity. The second part is based on a review of washback studies, such as Hughes' washback model, 15 Washback Hypotheses, and Green's concept of washback variability. Table 1 describes the 23 items listed in the student questionnaire. The Cronbach Alpha reliability for the 23 items on a 5-point Likert scale is 0.83. The 24<sup>th</sup> item asked which assessment did the students favor the most.

Table 1

#### *Items on the Classroom Assessment Questionnaire*

Part I	Items
Congruence with Planned Learning	1. How I am assessed and what I am assessed upon relate to what is done in the English class. 2. What is taught in the English class can help me to prepare for the assessments.
Authenticity	1. I have learned useful knowledge from the assessment. 2. I find the English assessment tasks to be relevant to the real world. 3. I find the English assessment tasks to be relevant to the future workplace.
Student Consultation	1. I am aware of how I will be assessed and what I will be assessed upon in the English class. 2. My teacher has explained to me the purpose of each assessment.

Transparency	<ol style="list-style-type: none"> <li>1. I understand how the English assessment tasks will be marked.</li> <li>2. I am told in advance when I will be assessed.</li> <li>3. I know what is needed to successfully accomplish the English assessment and get high marks.</li> </ol>
Diversity	<ol style="list-style-type: none"> <li>1. I am given a choice of assessment tasks that suit my ability.</li> <li>2. I am allowed to complete assessment tasks at my own speed.</li> </ol>
Part II	Items
Washback (a) Test demands (content, format, and complexity) (b) Purpose of the test (c) Stakes of the test (d) Test difficulty (e) Test preparation resources (f) Teaching methods (g) Learning content	<ol style="list-style-type: none"> <li>1. The assessment is important to me.</li> <li>2. The assessment is easy to me.</li> <li>3. The assessment can measure my English ability.</li> <li>4. Preparing for the assessment can enhance my motivation to learn English.</li> <li>5. Preparing for the assessment has enhanced my four English skills (listening, reading, writing, and speaking).</li> <li>6. Preparing for the assessment has enhanced my English proficiency levels.</li> <li>7. I have experienced a variety of learning methods and activities while preparing for the assessment.</li> <li>8. Preparing for the assessment is a meaningful and worthwhile experience.</li> <li>9. I have spent a lot of time preparing for the assessment.</li> <li>10. I am under pressure while preparing for/taking the assessment.</li> <li>11. I am in favor of this assessment.</li> </ol>
Part III	Items
Favorite classroom assessment	<ol style="list-style-type: none"> <li>1. Traditional pen-and-pencil test.</li> <li>2. Individual presentation.</li> <li>3. Group project.</li> <li>4. A combination of the above three.</li> </ol>

### Data collection procedure

The study gave clear instructions over three assessments, including 1) two traditional multiple-choice listening and reading examinations (summative assessment in Tables 2 & 3), 2) a two-minute presentation (formative assessment 1 in Tables 2 & 3), and 3) group audio PowerPoint presentation (formative assessment 2 in Tables 2 & 3), as to their purpose, test tasks, test format, test time, marking criteria, and preparation. For the two formative assessments, the students were given oral feedback and written feedback during the process of their presentation.

After the students completed the three assessments, the classroom assessment questionnaires were distributed in class. Each student was rewarded with a free glass of fruit juice to thank them on their effort in answering the questionnaires.

### Data analysis

IBM SPSS Statistics 22 was used to analyze the quantitative data. We utilized descriptive statistics to calculate the frequency distributions, means, and standard deviations of the questionnaire items and test scores. We also used inferential statistics (e.g. independent t-test, one-way ANOVA) to check for a statistical significance level of .05 ( $p < 0.05$ ). The effect size is also reported to help readers

understand the magnitude of the significant differences found (Larson-Hall 2012; Larson-Hall & Plonsky, 2015), where .10, .30, and .50 are the benchmarks for the small, medium, and large effects in Cohen's (1988) effect size interpretation system.

## **Results & discussion**

### **1. What are the differences between students' perceptions of summative and formative assessments?**

According to one-way ANOVA analysis, we present the main findings that illustrate the statistically significant differences and effect sizes between students' perceptions of these three types of assessments, as shown in Tables 2 & 3.

In terms of consultation transparency, students are in favor of summative assessment more than the other types to a small degree. The students have a slightly greater understanding of the purpose and the marking scale of the summative assessment, possibly due to the fact that they have primarily taken summative written exams during the course of their studies. In terms of transparency, students have a greater understanding of how to prepare for the formative oral assessment than the other two assessments between a small and medium degree. This finding is likely, because there is only one oral question for the formative oral assessment, and students can devote their full attention to preparing for that single question. However, there is a lot more material covered in the summative and formative cooperative assessments, and so what students have prepared for may not even appear on the test, and accordingly, students have spent much more time completing these two kinds of assessments. In terms of diversity and authenticity, students favor formative assessments more than summative assessments to a medium degree. In their opinion, they can choose the assessment tasks that suit their ability at their speed. In addition, students contend that they can learn useful knowledge relevant to the real world from either the formative oral assessment or formative cooperative assessment.

In terms of washback effects, students favor summative assessments to a medium degree, because they are fair, can measure students' abilities, and help them to improve their vocabulary, listening, and reading skills. However, students favor formative assessments because they think



preparing for these two types of assessments are a meaningful and worthwhile learning experience. They stated that they had improved their writing and oral skills from formative assessments. Interestingly, although students stated that formative oral assessments increased the amount of pressure they experienced, they prefer this type of assessment than the summative assessment to a medium degree.

When asked which type of classroom assessments they preferred to take, 62% of the participants stated they liked the combination of the traditional pen-and-pencil test, individual presentation, and group project. Another 10-15% of the participants preferred each of the three classroom assessments respectively.

Based on the aforementioned findings, the preparation for summative assessments elicited a greater degree of test-oriented learning, whereas formative assessments enhanced students' motivation to learn English. Students have favorable opinions of different assessments, because of the positive effects they brought about on students' learning. Therefore, it is believed that an appropriate combination of summative and formative assessment tasks is beneficial for their learning.

Table 2  
*Students' perceptions of the three types of assessments (1)*

Category	Questionnaire items	Assessment type (I)	Assessment type (J)	Mean (I-J)	Standard error	Sig.	r
<b>Student consultation</b>	My teacher has explained to me the purpose of each assessment.	Summative	Formative 2	.233	.083	.015	.23
		Formative 1	Formative 2	.289	.083	.002	.31
<b>Transparency</b>	1 I understand how the English assessment tasks will be marked.	Summative	Formative 1	.308	.114	.020	.22
		Summative	Formative 2	.414	.114	.001	.31
	2 I know what is needed to successfully accomplish the English assessment and get high marks.	Formative 1	Summative	.346	.106	.003	.28
		Formative 1	Formative 2	.368	.106	.002	.30
<b>Diversity</b>	1 I am given a choice of assessment tasks that suit my ability.	Formative 2	Summative	.506	.116	.000	.37
	2 I am allowed to complete assessment tasks at my own speed.	Formative 2	Summative	.263	.108	.041	.20
<b>Authenticity</b>	1 I have learned useful knowledge from the assessment.	Formative 2	Summative	.291	.106	.018	.23
		Formative 2	Formative 1	.300	.106	.014	.24
	2 I find the English assessment tasks to be relevant to the real world.	Formative 1	Summative	.346	.121	.013	.24

Table 3

*Table 3: Students' perceptions of the three types of assessments (2)*

Washback: Questionnaire items	Assessment type (I)	Assessment type (J)	Mean (I-J)	Standard error	Sig.	r
1 The assessment is fair.	Summative	Formative 1	.243	.095	.030	.20
	Summative	Formative 2	.602	.095	.000	.37
	Formative 1	Formative 2	.359	.095	.001	.31
2 The assessment can measure my English ability.	Summative	Formative 2	.595	.114	.000	.37
	Formative 1	Formative 2	.427	.114	.001	.31
3 I am nervous and feel pressure when completing the assessment tasks.	Summative	Formative 1	-.421	.145	.011	.24
	Summative	Formative 2	.724	.146	.000	.37
	Formative 1	Formative 2	-1.14	.146	.000	.37
4 Preparing for the assessment is a meaningful and worthwhile experience.	Summative	Formative 2	-.310	.095	.003	.28
5 I have studied a variety of learning methods and activities from the assessment.	Summative	Formative 1	-.458	.118	.000	.37
	Summative	Formative 2	-.891	.119	.000	.37
	Formative 1	Formative 2	-.433	.119	.001	.31
6 Preparing for the assessment has increased my knowledge of grammar and bank size of vocabulary.	Summative	Formative 1	.280	.089	.005	.27
	Formative 1	Formative 2	-.329	.089	.001	.31
7 Preparing for the assessment has enhanced my listening skills.	Summative	Formative 1	.467	.121	.001	.31
	Summative	Formative 2	.528	.131	.000	.31
8 Preparing for the assessment has enhanced my reading skills.	Summative	Formative 1	.776	.107	.000	.37
	Summative	Formative 2	.384	.107	.001	.31
	Formative 1	Formative 2	-.392	.107	.001	.31
9 Preparing for the assessment has enhanced my oral skills.	Summative	Formative 1	-1.290	.123	.000	.37
	Summative	Formative 2	-1.330	.123	.000	.37
10 Preparing for the assessment has enhanced my writing skills.	Summative	Formative 2	-.622	.151	.000	.37
	Formative 1	Formative 2	-.791	.151	.000	.37
11 I am in favor of the assessment.	Summative	Formative 1	-.561	.133	.000	.37
	Summative	Formative 2	-.503	.134	.001	.31

## 2. What are the differences between students' perceptions of summative and formative assessments based on their levels of proficiency?

According to an independent t-test, we present the main findings regarding the statistically significant differences and effect sizes between high- and low-proficiency students' perceptions of the three types of assessments, as shown in Tables 4-7.

High-proficiency students favor summative assessments more than low-proficiency students because they can increase the size of their vocabulary bank. Low-proficiency students favor summative assessments, only because these are important tests for them, and they might need to repeat the class if they cannot pass the summative assessments.

High-proficiency students favor formative assessments more than low-proficiency students because these tests can enhance their vocabulary, and they are more aware of the assessment tasks in terms of how to manage their time to prepare for them and when they will take place. Low-proficiency students favor formative assessments more than high-proficiency students because they can measure their English ability, enhance their listening and reading skills, and believe that these assessment tasks are relevant to real life. These findings appear to indicate that high-achieving students are in favor of the summative assessment due to its congruence with planned learning and transparency; most of these students admitted that they learned more from preparing for the summative assessment. Furthermore, low-achieving students viewed the cooperative group assessment positively because of its diversity.

The effect sizes for the significant differences between high-proficiency and low-proficiency students in terms of their preferences in summative and formative assessments are to a small degree. This finding indicates that the type of assessment may not be the major concern for students, but how students view the stakes, purpose, task demands, and difficulty of the assessment will influence their learning effects and learning attitude. Given these findings, low-proficiency students seem to benefit more from formative assessments, because they believe these assessments can measure their English proficiency, and preparing for formative assessments can improve their listening and reading skills. However, these formative assessments did not produce a significant improvement in their learning outcomes. Traditional paper-and-pencil written assessments may be beneficial for high-proficiency students to increase the size of their vocabulary bank. However, in order to motivate low-proficiency students, formative oral and cooperative assessments may be another solution for teachers in the English classroom.

As suggested by Peterson & Siadat (2009) and O'Neill (2102), this study echoes the above findings to propose the type of balanced assessment in which the teacher brings many and various strands of assessment together in a coherent way that addresses the desired goals and takes account of opportunities and constraints in the setting concerned. No matter whether a summative or formative assessment is practiced in class, it is essential to provide congruence with planned learning, authenticity, student consultation, transparency, diversity, and intended positive washback effects, such as meaningful learning experience, enhanced motivation, and proficiency levels.

High-achieving and low-achieving students, based on the findings, have different preferences toward the summative and formative assessments regarding their washback effects. In particular, low-achieving students focused on the stakes, the demand, and the purposes of the assessment. In contrast, high-achieving students paid attention to the quality of classroom

assessment, such as its student consultation, transparency, and diversity, in addition to washback effects. Given these findings, when implementing both formative and summative assessments, teachers should clearly inform their students how, what, and when they will be assessed in order for them to know what they should prepare or work on for the assessment with the final goal of helping teachers understand their students' learning processes and outcomes.

Another finding that should catch teachers' attention is that students have more favorable perceptions of formative assessment over summative assessment. This finding appears to promote the concept of assessment for learning (Klenowski, 2009), where teachers design assessment tasks that require students to do what they are interested in by using a variety of classroom activities, so that teachers can use the data collected from the students to help them improve students' learning. In Taiwan's education, university teachers usually instruct a large class size (around 45-60 for general English classes), and therefore summative assessments are usually given to assist teachers in understanding their students' learning outcomes, because it is easier to mark summative assessments. Therefore, time and labor constraints may hinder teachers' use of formative assessments. It is thus suggested that teachers be provided with more educational resources to encourage them to use formative assessments for understanding students' learning process and to help improve their learning as the end goal.

Table 4

*High- and low-proficiency students' perceptions of the three types of assessments (1)*

High-proficiency learners' perceptions of summative written assessment				Low-proficiency learners' perceptions of summative written assessment			
Ranking	Items	M	SD	Ranking	items	M	SD
Top 1	I am told in advance when I will be assessed.	4.63	.486	Top 1	I am told in advance when I will be assessed.	4.47	.573
Top 2	My teacher has explained to me the purpose of the assessment.	4.37	.561	Top 2	Passing the assessment is important to me.	4.47	.573
Top 3	Preparing for the assessment has increased my knowledge of grammar and amount of vocabulary.	4.35	.590	Top 3	Preparing for the assessment has enhanced my reading skills.	4.29	.658
Top 4	Preparing for the assessment has enhanced my reading skills.	4.35	.653	Top 4	My teacher has explained to me the purpose of the assessment.	4.20	.524
Top 5	I understand how the assessment is evaluated.	4.21	.776	Top 5	How the assessment is evaluated is fair.	4.20	.558
Btm 1	Preparing for the assessment has enhanced my written skills.	2.87	1.253	Btm 1	Preparing for the assessment has enhanced my oral skills.	2.95	1.161
Btm 2	Preparing for the assessment has enhanced my oral skills.	2.92	.837	Btm 2	Preparing for the assessment has enhanced my written skills.	3.05	1.096
Btm 3	I am nervous and feel pressure when taking the assessment.	3.25	1.064	Btm 3	I have spent a lot of time preparing for the assessment.	3.29	.956
Btm 4	I am in favor of the assessment.	3.25	1.046	Btm 4	I am nervous and feel pressure when taking the assessment.	3.29	1.100
Btm 5	I have experienced a variety of	3.44	1.037	Btm 5	I am in favor of the assessment.	3.33	1.055

learning methods and activities while preparing for the assessment.						
---	--	--	--	--	--	--

Note: summative= written assessment; formative 1 = individual oral presentation; formative 2 = group project.

Table 5

*High- and low-proficiency students' perceptions of the three types of assessments (2)*

High-proficiency learners' perceptions of formative assessment (1)				Low-proficiency learners' perceptions of Formative assessment (1)			
Ranking	Items	M	SD	Ranking	items	M	SD
Top 1	I am told in advance when I will be assessed.	4.52	.641	Top 1	Passing the assessment is important to me.	4.44	.536
Top 2	My teacher has explained to me the purpose of the assessment.	4.37	.595	Top 2	My teacher has explained to me the purpose of the assessment.	4.31	.540
Top 3	I am aware of how I will be assessed and what I will be assessed upon in the English class.	4.35	.556	Top 3	I am told in advance when I will be assessed.	4.31	.635
Top 4	Preparing for the assessment has enhanced my oral skills.	4.23	.703	Top 4	Preparing for the assessment has enhanced my oral skills.	4.29	.875
Top 5	Passing the assessment is important to me.	4.19	.817	Top 5	I am in favor of the assessment.	4.25	.700
Btm 1	Preparing for the assessment has enhanced my written skills.	2.63	1.138	Btm 1	Preparing for the assessment has enhanced my written skills.	2.95	1.113
Btm 2	I have spent a lot of time preparing for the assessment.	3.21	1.016	Btm 2	I have spent a lot of time preparing for the assessment.	3.53	.979
Btm 3	Preparing for the assessment has enhanced my listening skills.	3.33	1.024	Btm 3	Preparing for the assessment has enhanced my reading skills.	3.58	.875
Btm 4	Preparing for the assessment has enhanced my reading skills.	3.50	1.038	Btm 4	I am nervous and feel pressure when taking the assessment.	3.73	1.079
Btm 5	The assessment can measure my English ability.	3.54	.753	Btm 5	Preparing for the assessment has enhanced my listening skills.	3.75	.947

Table 6

*High- and low-proficiency students' perceptions of the three types of assessments (3)*

High-proficiency learners' perceptions of formative assessment (2)				Low-proficiency learners' perceptions of formative assessment (2)			
Ranking	Items	M	SD	Ranking	Items	M	SD
Top 1	I am told in advance when I will be assessed.	4.65	.480	Top 1	I have experienced a variety of learning methods and activities while preparing for the assessment.	4.44	.536
Top 2	Preparing for the assessment has increased my knowledge of grammar and amount of vocabulary.	4.33	.550	Top 2	I am told in advance when I will be assessed.	4.31	.540
Top 3	I have experienced a variety of learning methods and activities while preparing for the assessment.	4.29	.776	Top 3	Preparing for the assessment has enhanced my oral skills.	4.31	.635
Top 4	Preparing for the assessment has enhanced my oral skills.	4.29	.723	Top 4	Passing the assessment is important to me.	4.29	.875
Top 5	I am aware of how I will be assessed and what I will be assessed upon in the English class.	4.27	.717	Top 5	Preparing for the assessment is a meaningful and worthwhile experience.	4.25	.700

Note: summative= written assessment; formative 1 = individual oral presentation; formative 2 = group project.

Table 6

*(Continued)*

High-proficiency learners' perceptions of formative assessment (2)				Low-proficiency learners' perceptions of formative assessment (2)			
Ranking	Items	M	SD	Ranking	items	M	SD
Btm 1	I am nervous and feel pressure when taking the assessment.	2.62	1.105	Btm 1	I am nervous and feel pressure when taking the assessment.	2.95	1.113
Btm 2	The assessment can measure my English ability.	3.21	.977	Btm 2	The assessment can measure my English ability.	3.53	.979
Btm 3	Preparing for the assessment has enhanced my listening skills.	3.27	.931	Btm 3	How the assessment is evaluated is fair.	3.58	.875
Btm 4	I find the English assessment tasks to be relevant to the real world.	3.50	1.038	Btm 4	I have spent a lot of time preparing for the assessment.	3.73	1.079
Btm 5	Preparing for the assessment has enhanced my listening skills.	3.54	.979	Btm 5	Preparing for the assessment has enhanced my writing skills.	3.75	.947

Table 7

*A summary of high- and low-proficiency students' perceptions of the three types of assessments (4)*

A comparison of high-proficiency learners' and low-proficiency learners' perceptions of the three types of assessments

Assessment type	Items	<i>t</i>	<i>Sig.</i>	<i>r</i>
Summative	Passing the assessment is important to me.	1.989	.049	.19
	I am allowed to complete assessment tasks at my own speed.	2.347	.021	.22
	Preparing for the assessment has increased my knowledge of grammar and amount of vocabulary.	2.432	.017	.23
Formative 1	The assessment can measure my English ability.	2.690	.008	.25
	Preparing for the assessment has increased my knowledge of grammar and amount of vocabulary.	2.069	.041	.20
	Preparing for the assessment has enhanced my listening skills.	2.197	.030	.21
	I find the English assessment tasks to be relevant to the real world.	2.450	.016	.23
	I am in favor of the assessment.	2.017	.046	.19
Formative 2	Passing the assessment is important to me.	2.003	.048	.19
	I am aware of how I will be assessed and what I will be assessed upon in the English class.	2.198	.030	.21
	I am told in advance when I will be assessed.	2.668	.009	.25
	Preparing for the assessment has enhanced my listening skills.	2.122	.036	.20
	I find the English assessment tasks to be relevant to the real world.	2.489	.014	.24

Note: summative= written assessment; formative 1 = individual oral presentation; formative 2 = group project.

### Conclusions

The purpose of this study has been to explore Taiwan university students' perceptions of summative and formative classroom assessment tasks in their English classes. To address this issue, a questionnaire was designed, based upon Dorman and Knightley's (2006) PATI and Green's (2007) model of washback. According to survey questionnaires given to 107 first-year

undergraduate students at one Taiwan university, this study discovers that: 1) students were in favor of the summative assessment due to its congruence with planned learning and transparency, with most students admitting that they learned more from preparing for the summative assessment; 2) students viewed the cooperative group assessment positively, because of its diversity; 3) preparation for summative assessments elicited a greater degree of test-oriented learning for respective skills, whereas formative assessments enhanced students' motivation to learn English for productive skills; and 4) students believed that an appropriate combination of summative and formative assessment tasks was beneficial for their learning. Based on the findings, a combination of summative and formative assessments should be given appropriately to better understand students' learning outcomes and learning processes. Since the findings were drawn from a small sample size, further research should recruit a larger size of student participants for better generalizing the research findings. Teachers, who are also the major stakeholder of the assessments, should also be queried in order to present additional evidence for understanding the quality and washback of classroom assessment.



### References

- Alderson, C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115-129.
- Black, P. (2000). Research and the development of education assessment. *Oxford Review of Education*, 26(3&4), 407-419.
- Brown, H. D. (2004). *Language assessment: principles and classroom practices*. New York: Pearson Education/Noyce.
- Cheng, L. (2005). *Changing language teaching through language testing: washback study*. Cambridge: Cambridge University Press.
- Cheng, L. (2014). Consequences, impact, and washback. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 1130–1146). John Wiley & Sons.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2<sup>nd</sup> ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Dorman, J. P., & Knightley, W. M. (2006). Development and Validation of an Instrument to Assess Secondary School Students' Perceptions of Assessment Tasks. *Educational Studies*, 32(1), 47-58.
- Earl, L. (2003). *Assessment as learning: Using classroom to maximize student learning*. Thousand Oaks, CA: Corwin Press.
- Earl, L., & Katz, S. (2005). *Rethinking Classroom Assessment With Purpose in Mind*. Winnipeg, Manitoba: Western Northern Canadian Protocol.
- Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 191-210). Mahwah, New Jersey: Lawrence Erlbaum.
- Gan, Z. (2009). IELTS Preparation Course and Student IELTS Performance: A Case Study in Hong Kong. *RELC Journal*, 40(1) 23-41.
- Gebril, A., & Brown, G. (2014). The effect of high-stakes examination systems on teacher beliefs: Egyptian teachers' conceptions of assessment. *Assessment in Education: Principles, Policy & Practice*, 21(1), 16-33.
- Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge: Cambridge University Press.
- Hayes, B., & Read, J. (2003). IELTS Test preparation in New Zealand: Preparing students for the IELTS academic module. In R. Tolloh (Ed.), *IELTS Research Report 4* (pp. 153-206). Canberra: IELTS Australia Pty Limited.
- Hughes, A. (1993). *Backwash and TOEFL 2000*. Unpublished manuscript, University of Reading, England.
- Hughes, A. (2003). *Testing for language teachers* (2<sup>nd</sup> ed.). Cambridge: Cambridge University Press.
- Klenowski, V. (2009) Assessment for learning revisited: an Asia-Pacific perspective. *Assessment in Education: Principles, Policy and Practice*, 16(3), 263-268.

- Larson-Hall, J. (2012). Our statistical intuitions may be misleading us: Why we need robust statistics. *Language Teaching*, 45(4), 460-474.
- Larson-Hall, J., & Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning*, 65(S1), 127-159.
- O'Neill, G. (2012). UCD Assessment ReDesign Project: The Balance Between Assessment FOR and OF Learning, available at <https://www.ucd.ie/t4cms/UCDTLA0044.pdf>
- Peterson, E., & Siadat, M. V. (2009). Combination of Formative and Summative Assessment Instruments in Elementary Algebra Classes: A Prescription for Success. *Journal of Applied Research in the Community College*, 16(2), 92-102.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22(2), 142-173.
- Qi, L. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. *Assessment in Education*, 14(1), 51-74.
- Shih, C. (2007). A new washback model of students' learning. *The Canadian Modern Language Review*, 64(1), 135-162.
- Smyth, E., & Banks, J. (2012). High stakes testing and student perspectives on teaching and learning in the Republic of Ireland. *Education Assessment Evaluation and Accountability*, 24:283-306. DOI 10.1007/s11092-012-9154-6.
- Tsai, Y., & Tsou, C. (2009). A standardized English Language Proficiency test as the graduation benchmark: student perspectives on its application in higher education. *Assessment in Education: Principles, Policy & Practice*, 16(3), 319-330.
- Wall, D., & Horák, T. (2006). *The impact of changes in the TOEFL® examination on teaching: Phase 1, The baseline study* (TOEFL-MS-34). Princeton, NJ: ETS.
- Wall, D., & Horák, T. (2008). *The impact of changes in the TOEFL® examination on teaching and learning in Central and Eastern Europe: Phase 2, Measuring change* (TOEFL iBT-05). Princeton, NJ: ETS,
- Wall, D., & Horák, T. (2011). *The Impact of Changes in the TOEFL® Exam on Teaching in a Sample of Countries in Europe: Phase 3, The Role of the Coursebook. Phase 4, Describing Change* (TOEFL iBT-17). Princeton, NJ: ETS.
- Xie, Q. (2013). Does test preparation work? Implications for score validity. *Language Assessment Quarterly*, 10, 196-218.
- Xie, Q. (2015). Do component weighting and testing method affect time management and approaches to test preparation? A study on the washback mechanism. *System*, 50, 56-88.
- Zhan, Y., & Andrew, S. (2014). Washback effects from a high-stakes examination on out-of-class English learning: insights from possible self theories. *Assessment in Education: Principles, Policy & Practice*, 21(1), 71-89.

Appendix 1: Alderson and Wall's fifteen washback hypotheses  
(1993, pp. 120-121)

- 1) A test will influence teaching.
- 2) A test will influence learning.
- 3) A test will influence what teachers teach; and
- 4) A test will influence how teachers teach; and therefore by extension from (2) above:
- 5) A test will influence what learners learn;
- 6) A test will influence how learners learn.
- 7) A test will influence the rate and sequence of teaching; and
- 8) A test will influence the rate and sequence of learning.
- 9) A test will influence the degree and depth of teaching.
- 10) A test will influence the degree and depth of learning.
- 11) A test will influence attitudes to the content, method, etc. of teaching and learning.
- 12) Tests that have important consequences will have washback; conversely,
- 13) Tests that do not have important consequences will have no washback.
- 14) Tests will have washback on all learners and teachers.
- 15) Tests will have washback effects for some learners and some teachers, but not for others.

## A Comparative Study of Individual and Collaborative Oral Languageing for L2 Speech Act Production

**Marzieh Bagherkazemi\***

*Islamic Azad University (South Tehran Branch), Iran  
Department of English Language Teaching*

### Abstract

The role of languageing in second language acquisition (SLA) has been widely investigated since its postulation in the 1980s, though only a few studies have addressed languageing in speech act production as an aspect of second language (L2) pragmatic development. The present study was designed to compare the nature of languageing produced by 45 intermediate English as a Foreign Language (EFL) learners while completing five apology and five request written discourse completion tasks (WDCTs) individually (N=15) and collaboratively (in pairs) (N=30; 15 pairs). Following a two and a half-hour workshop, individual think-aloud protocols (totalling 304 minutes) and paired interactions (totalling 392 minutes) were transcribed. Subsequently, episodes of noticing, reflection and hypothesis testing, as the three main functions of languageing, were detected in the transcripts. They were coded by two coders based on a coding scheme specifically designed with reference to speech act production. The coded episodes were then subjected to qualitative comparisons. In general, the analyses revealed the greater potential of collaborative languageing to induce the noticing of more social context variables (SCVs) involved in performance. Collaborative languageing also nested comparative and more profound reflections, as well as successful output modifications following the greater number of hypothesis testing episodes it led to. The findings are discussed in light of the sociocultural notions of inter-psychological learning mechanisms involved in collaborative dialoguing.

**Keywords:** collaborative dialoguing; hypothesis testing; languageing; noticing; output; pragmatics; reflection; speech act.

---

\* *m\_bagherkazemi@azad.ac.ir*

## Introduction

Studies on the role of learners' individual and collaborative language productions in SLA gained momentum in the mid-1980s pioneered by Swain's Comprehensible Output Hypothesis (COH) after more than a decade of majorly input-oriented SLA research. Before the postulation of the COH, studies on learner language were mainly intended to feedback into the input provided to the learners, and as such failed to recognize the ways in which such productions could assist the language learning process. Put forth in reaction to Krashen's (1985) Comprehensible Input Hypothesis, Swain's (1985) COH hinges on the significance of learners' attempts at language production for their language acquisition.

The essentiality of output production – later termed as “*linguaging*” (Swain, 2006) following the subscription of COH to Vygotsky's sociocultural theory – for SLA lies in its potential to (a) facilitate noticing of L2 features, (b) induce metalinguistic reflections, and (c) lead to the generation and testing of language-related hypotheses. Based on the sociocultural account of language development, language is an essential tool which mediates and regulates the socially-situated process of language learning. This process involves learners' internalization of language knowledge co-constructed in the course of their interaction with physical artifacts and/or more capable others (Lantolf, 2011; Swain & Lapkin, 2001).

Linguaging – whether individual as “*private speech*” or collaborative as “*collaborative dialoguing*” in sociocultural terms - has been mainly investigated in terms of its potential for language learning (e.g., Bao, 2019; Brooks, et al., 2010; Ishikawa, 2013, 2015; Ishikawa & Suzuki, 2016; Jia, 2015; Knouzi, Swain, Lapkin, & Brooks, 2010; Li, 2015; Liang, 2014; Moradian, Miri, & Hossein Nasab, 2017; Suzuki, 2009, 2012, 2017; Suzuki & Itagaki, 2009). Such research has primarily targeted grammar and writing, and fallen short of adequately addressing other language skills and components, though findings allude to the general effectiveness of linguaging. An area ripe for research is the potential contribution of individual and collaborative linguaging to the learning of L2 pragmatic features, speech acts included, in terms of Swain's (1985) postulated functions of learner output: noticing, metalinguistic reflection, and hypothesis testing.

Since the inclusion of discourse and sociolinguistic competencies in models of linguistic and communicative competence in the 1980s (e.g., Canale & Swain, 1980), interlanguage pragmatics (ILP) has been the focus of much SLA research, gaining momentum at the turn of the

21st century. Among all pragmatic features, speech acts have been targeted the most in such research (see Taguchi, 2011, 2015), owing to their cross-culturally and cross-linguistically variant realizations (Bardovi-Harlig, 2001). Instructional pragmatics, though mainly concerned with the implicit/explicit distinction, rooted in Schmidt's (1993) Noticing Hypothesis as a cognitive take on SLA, in the first decade of ILP research, is now being increasingly studied within more interactionist theoretical frameworks (e.g., Tajeddin & Tayebipour, 2012). There is some research evidence as to the greater efficacy of collaborative output for the production of speech acts as indicated in WDCT performance; however, the nature of languaging learners engage in individually or in collaboration with peers in relation to the three hypothesized functions of L2 output, namely noticing, metalinguistic reflection and hypothesis testing, has not been qualitatively investigated. The present study was designed to shed light on differences between learners' individual and collaborative languaging while completing WDCTs in terms of their potential to enhance these three L2 output functions.

### **Literature Review**

#### **ILP Development and Speech Act Production**

Pragmatics has been generally defined as the study of language in context (Bardovi-Harlig, 2001). In SLA research, the study of learners' mastery of and control over L2 pragmatic features has been referred to as ILP (Kasper & Rose, 2001). Since its uptake at the turn of the 21st century, ILP research can be characterized by two consecutive foci. Earlier research was almost exclusively devoted to the investigation of cross-culturally and cross-linguistically different realizations of pragmatic features; speech acts being the most frequently targeted feature. The findings generally evidenced sociopragmatic and pragmalinguistic differences in the way different languages' pragmatic features are verbalized as well as the dire consequences of breaching L2 pragmatic norms for the flow and success of communication. Studies of the sort fed into later research placing a premium on the teachability of pragmatic features (and the desirability of doing so) and subsequently sound instructional pragmatics approaches (Rose & Kasper, 2001; Taguchi, 2011, 2015).

Research into how best to teach L2 pragmatic features was in its early days largely reliant

on the Noticing Hypothesis (Schmidt, 1993). In pragmatic terms, this hypothesis would translate into the explicit provision of pragmalinguistic and sociopragmatic norms of performing specific L2 pragmatic features, or otherwise having L2 learners induce them from relevant teacher-provided input. This accounts for the surge of studies investigating implicit and (inductive and deductive) explicit instructional approaches in separate and comparative designs, most of which substantiated the superiority of the latter (see Taguchi, 2011, 2015). It is less than 10 years since other theoretical frameworks have been brought to bear on the instruction of speech acts and other pragmatic features. Related studies have been conducted on the basis of several accounts of SLA, including VanPatten's (1996) Input Processing Theory and Processing Instruction (e.g., Takimoto, 2009, 2010), Skill Acquisition theories (e.g., Li, 2012), Swain's (1985) COH (e.g., Tajeddin & Bagherkazemi, 2014; Jernigan, 2007), and Vygotsky's Sociocultural Theory (SCT) (e.g., Khatib & Ahmadi Safa, 2011; Ohta, 2005; Tajeddin & Tayebipour, 2012). The last approach to explaining SLA, to which COH also subscribed despite its postulated cognitive underpinnings in its early days, hinges on learners' own mediated language production for its acquisitional significance. According to Lantolf (2011), individual and collaborative language productions have the potential for mediating language acquisition and facilitating learners' self-regulated learning.

In the realm of ILP, studies on learners' output have fallen short of addressing its role in SLA, be it individual or collaborative either in separate or comparative designs, but rather as an outcome of instruction (Norouzian & Eslami, 2016). Moreover, those which have been carried out within an SCT framework have essentially addressed collaborative dialoguing, the major question being whether learners' interaction with expert L2 interactants (mainly native speakers) can help them move forward in their Zone of Proximal Development, as far as speech act performance is concerned (e.g., Niu, 2017; Tajeddin & Tayebipour, 2014). The potential of co-equal peers' scaffolding for ILP development was compared with that of expert/non-expert peers' scaffolding in Khatib & Ahmadi Safa's (2011) study. The results showed the greater benefits of the latter, but also the significant effect of the former on learners' speech act performance. The nature of languaging learners engaged in was not, however, investigated in this study. Within this SCT-grounded ILP research context, the present study was carried out to qualitatively compare the mediating role of learners' attempts at producing L2 speech acts (apologies and requests in the present study). Mediation was defined with reference to the functions attributed to languaging in

COH: noticing, reflection and hypothesis testing (see Procedure). The concept of languaging with its variants and status in SLA and language teaching research is sketched in the following section.

### **Languaging in SLA Research**

By definition, the term “languaging” refers to “the process of meaning making and shaping knowledge and experience through language” (Swain, 2006, p. 98). This definition resonates with the role assigned to language as a learning mediator in the sociocultural account of language development. Individual and collaborative languaging (“private speech” and “collaborative dialoguing,” respectively, in sociocultural terms) is, according to Swain (1985, 2006), an indispensable aspect of the language learning process. Private speech is defined by Negueruela and Lantolf (2006, p. 86) as “the intentional use of overt self-directed speech to explain concepts to the self,” whereas in collaborative dialoguing, “speakers are engaged in problem-solving and knowledge building” (Swain and Lapkin, 1998, p. 102). Languaging, whether individual or collaborative, can be either task-induced or teacher-imposed. In their review of studies into languaging and learner output, Niu and Li (2017) see the former as incidental and the latter as more conducive to learning.

Languaging could also be characterized in terms of its modality, as either oral or written. Studies on oral languaging have mainly targeted L2 learners’ grammar development, while written languaging has been investigated in relation to corrective feedback and translations, in addition to grammar explanations. The short-term benefits of written languaging for grammar accuracy and writing quality is evidenced in Moradian et al.’s (2017) study of Iranian EFL learners, and its long-term benefits for writing and lexical and grammar accuracy in Jia’s (2015) investigation with Chinese EFL learners. With regard to the quality of written languaging, Suzuki (2017) found both “languaging as noticing-only” and “languaging as noticing with reasons” beneficial for Chinese EFL learners’ writing accuracy development. The results led him to assign languaging a mediating problem-solving role.

Relevant to the concerns of the present study, research in the former category (Brooks et al., 2010; Knouzi et al., 2010; Li, 2015) has uniformly demonstrated the efficacy of oral languaging for the learning of grammar and resolution of cognitive conflicts, though this effectiveness is



mediated by the amount and quality of languaging, learners' prior knowledge of the targeted feature, and their proficiency level in collaborative dyads/groups; however, studies addressing oral languaging (a) in relation to other language components, (b) vis-à-vis traditional approaches to grammar instruction in comparative designs, or (c) produced by individual learners or as joint attempts with the variety of pairing/grouping options are yet to be carried out. Against this backdrop, the present study extended languaging research to ILP development, and more specifically to EFL learners' individual and paired attempts at completing apology and request WDCTs.

### **Methodology**

The present study was carried out to investigate the nature of EFL learners' individual and collaborative languaging while trying to produce the two speech acts of apology and request in 10 WDCTs. This section provides an account of the participants, instruments and procedure.

### **Participants**

For the purpose of the present study, a total of 45 intermediate female EFL learners (between 19 and 23 years of age) took part in the study. They were selected (from among an initial 58-member pool) through a convenience sampling procedure. They were all English language teaching (ELT) freshmen at Islamic Azad University (South Tehran Branch, Iran), had not resided in an English-speaking country, and belonged to four "Conversation" classes, as an obligatory course offered in the second semester of the study program. Following a workshop of two and a half hours (see Procedure), they were randomly assigned to two groups: an individual languaging group (ILG) and a collaborative or paired languaging group (CLG). The participants were homogenized in terms of their language proficiency and apology and request WDCT performance. These were controlled for as research has shown proficiency and initial knowledge of the learning target to determine the quantity and quality of languaging (see Jia, 2015).

## Instruments

The study involved two main instruments: the Quick Placement Test (QPT) and a 16-item WDCT. First, the participants were homogenized in terms of their language proficiency through the paper-and-pen version of the Quick Placement Test (QPT). This step was taken to warrant between-group comparisons in terms of aspects of the quality and quantity of noticing, reflection and hypothesis testing episodes. QPT is a widely used proficiency test developed conjointly by Cambridge ESOL Examinations Syndicate and Oxford University Press. It comprises 60 recognition-type cloze reading comprehension, vocabulary and grammar items in an ascending difficulty order, and its results can be reported along Association of Language Testers in Europe (ALTE) levels from “beginner” to “very advanced.” The test generally enjoys good validity and reliability (see Geranpayeh, 2003). In the present study, the participants scored between 36 and 45, and thus were designated as intermediate. The test took 35 minutes to complete, and a Cronbach’s Alpha Coefficient of .86 showed the reliability of the scores.

Second, as languaging research has found learners’ initial knowledge to play a role in the quantity and quality of languaging, the participants’ apology and request production ability were homogenized across the two groups through a 16-item WDCT test. The WDCTs were selected from among existing ones (e.g., Zand Moghaddam, 2012) in a way to represent (a) various combinations of power, distance, and imposition (as the three SCVs) implicating in speech act performance and (b) situations familiar in university life involving professor-student, student-student and student-parent role relationships. WDCT responses were all rated by a native speaker and the researcher (with an inter-rater correlation coefficient of .91) along a 6-point Likert scale in the tradition of Taguchi (2006), based on such concerns as comprehensibility, grammatical and discoursal felicity, as well as appropriacy. Prior to assigning the participants to ILG and CLG and following the workshop (see Procedure), the WDCT was administered. After ensuring the normality of WDCT scores (with skewness (-.55) and kurtosis (-1.22) values over their standard error estimates (.44 and .85, respectively) falling within the range of  $\pm 1.96$ ), those scoring within the range of one standard deviation (.19) from the mean (3.07) were assigned as the main participants (N=45).

## Procedure

The present study's implementation involved the following steps:

1. convenience sampling of the initial 58 participants;
2. administration of the QPT and the WDCT to control for general proficiency level as well as speech act production ability, and the inclusion of intermediate EFL learners scoring within one standard deviation from the mean of the WDCT scores (N=45);
3. offering a two and a half-hour workshop to all the 45 participants to (a) familiarize them with pragmatics, ILP, speech acts, and the three SCVs; (b) provide conversation-embedded apology and request samples produced by native speakers (five on apology and five on request) and analyze them based on these two speech acts' strategy sets; (c) model individual languaging (the instructor/researcher) and collaborative languaging (the instructor/researcher paired one of the workshop participants), drawing attention to the SCVs, politeness, and grammatical and discursal appropriateness.
4. random assignment of the participants into ILG (N=15) and CLG (N=30), and random pairing of CLG participants as co-equals;
5. having ILG and CLG participants complete 10 WDCTs (five on apology and five on request) through individual and collaborative (paired) languaging as modeled in the last phase of the workshop, and record their own voices and interactions using their cell phones' voice recorder application; and
6. coding the recordings, analysis of the quantity (frequency in the present study) and quality of noticing, reflection and hypothesis testing episodes, and comparison of the two sets of episodes (see Data Analysis Results for the operational definitions and the coding scheme).

## Data Analysis Results

The present study involved a comparison of individual and collaborative languaging in terms of the frequency of occurrence and nature of episodes of noticing, reflection, and hypothesis testing they induce. Answering this question involved the analysis of (a) think-aloud protocols of fifteen participants in ILG and (b) paired interactions of thirty participants in CLG, while completing 10

WDCTs (five on the speech act of apology and 5 on the speech act of request). The think-aloud protocols and paired interactions were audiotaped and transcribed. Transcribed data were then subjected to qualitative analyses in terms of Swain's (1985) three postulated functions of learner output, i.e. noticing, hypothesis testing, and reflection. Upon an initial examination of the data, operational definitions of the three functions of languaging were adopted (and minimally adapted) from an earlier work by the researcher on learner output (Bagherkazemi, 2014), borrowing ideas from Jernigan (2007), Shehadeh (2002), Swain (1995, 2006), and Swain and Lapkin (1998). Languaging functions were defined as follows in the present study:

1. **Noticing:** (a) the first implicit or explicit mention of power/status, distance/familiarity and/or imposition under various rubrics, either before or after uttering the speech act, and (b) showing awareness of gaps in one's sociopragmatic and pragmalinguistic knowledge due to failed attempts at producing the speech act in question, through either implicit admission (e.g., "How can I say it?") or explicit admission (e.g., "I don't know how to say it.") in individual languaging, and implicit or explicit request for information in collaborative languaging (e.g., A asks B, "I wonder if I can or I wonder if I could?");
2. **Reflection:** using language individually or collaboratively to contemplate the situation, politeness issues, the possible interaction/clash of the three SCVs with each other and with politeness (e.g., "I should be very polite in this situation."), as well as the appropriacy of certain speech act strategies and semantic formulae (e.g., "If I don't tell her the reason in my apology, she will get upset.");
3. **Hypothesis testing:** individual or collaborative trial-and-error episodes regarding conjectures about the correspondence of situation-specific SCVs and the expressed speech act strategies and semantic formula, based on either own internal feedback in individual languaging or external feedback provided by one's interlocutor in collaborative languaging, leading to output modifications. Table 1 shows the descriptive codes along with their descriptors.

Table 1

*Individual and Collaborative Linguaging Coding Scheme* (reproduced from Bagherkazemi, 2014, p. 185)

Languaging function	Descriptive code	Descriptor
Noticing	N1	Learners notice one SCV.
	N2	Learners notice two SCVs.
	N3	Learners notice three SCVs.
	NKG	Learners notice sociopragmatic and/or pragmalinguistic knowledge gap.
Hypothesis testing	HT	Learners engage in trial-and-error episodes regarding sociopragmatic–pragmalinguistic mappings.
Reflection	R	Learners contemplate the situation, SCVs and/or politeness.

*Note.* SCV= social context variable (power, distance, imposition).

After the development of the coding scheme, the data were coded twice: once by the researcher (Coder 1) and once by a 36 year-old female university instructor (Coder 2). Coder 2 held a Ph.D. in English Language Teaching, and had a 10-year teaching experience at different language schools and universities in Iran. In advance of coding, she was briefed on the three functions of languaging and their operational definitions for the purpose of the study as well as the coding scheme. Subsequently, the two coders' codings were compared for the purpose of locating and discussing the discrepancies. Instances of noticing, hypothesis testing, and reflection not detected by one of the coders (N=13), and those coded differently by the two coders (N=18) were discussed, and agreements reached. Ambiguity lay in instances of *noticing* contained in the learners' *reflections* over the situation, as in Example 1.

#### **Example 1:**

**Apology situation:** You are a teacher. You promised one of your students to bring him/her a book on Wednesday afternoon, but you forgot. The student waited for you at the door of your office for one hour. Today is Thursday, and the student comes to your office again; you apologize to him/her.

**Think-aloud protocol:** *I think it is students who should ask their teacher for help, so it's ok to ask for something for a second or third time as a student.* It doesn't matter much, so teachers do not usually feel ashamed in these situations. As the teacher, I would say, "Sorry I forgot it yesterday as I had a busy schedule this week, but you can have it now."

The coders agreed that the italicized section of the think-aloud protocol in Example 1 be taken as an instance of "reflection" over the situation, and the two underlined parts be counted as instances of *noticing* of "power" and "imposition," respectively. Table 2 presents the finalized frequency of occurrence of instances of noticing, reflection, and hypothesis testing detected in the data.

Table 2

*Frequencies of Noticing, Reflection, and Hypothesis Testing in Individual/Collaborative Languageing*

Languageing function		Frequency of occurrence	
		Individual languageing	Collaborative languageing
Noticing	N1	72	24
	N2	90	103
	N3	3	54
	NKG	13	13
Hypothesis testing(HT)		12	30
Reflection (R)		78	80

*Note.* N1= Learners notice one SCV; N2= Learners notice 2 SCVs; N3= Learners notice 3 SCVs (where SCV= social context variables of power, distance, and imposition); NKG= Learners notice pragmatic knowledge gaps.

### Noticing in Individual and Collaborative Languageing

A comparison of individual and collaborative languageing in terms of the SCVs indicated that a larger number of such variables were generally noticed for each situation in paired interactions than in individual languageing. Participants in ILG showed consciousness of those SCVs which seemed to have clearer implications for verbalizing the speech act: They made a mention of one or two SCVs in the majority of cases (N=162) but failed to notice all three SCVs together for more than three situations. In Example 1, the learner has implicitly referred to *power* and the *low imposition* involved in producing the speech act, i.e. the low face threat likely to incur on the student. Failure to notice one or more of the SCVs did in some cases mislead the ILG participants

in their choice of appropriate speech act strategies and semantic formula. In Example 2, the learner has failed to notice “distance,” despite its being explicitly mentioned in the situation prompt, and this failure seems to have caused her to hesitate over the appropriate address term. Had she noticed this SCV, the choice would have been more easily made.

**Example 2:**

**Request situation:** You are doing your research project, and need to interview the president of your university. The president was your teacher, and you know him quite well. You know he is very busy and has a tight schedule. You still want to ask the president to spare one or two hours for your interview.

**Think-aloud protocol:** [reads the situation] So difficult, so tough! the president of the university, he is serious; so I should go and make a request ... request in a very ... very polite way, because he is in a position that is very ... maybe ...tough, although I know him well. On the other hand, I have no other way... I have to go. I must go because it is an important project. What can I say? How should I start “Mr. President,” or should I use his name? ... “Mr. Amini,” for example, “I was wondering if you kindly gave me some time for an interview.” I’m not sure if it is polite and formal enough.

On the other hand, participants in CLG noticed more than one SCV in the majority of cases; interactions in which two SCVs were noticed were paramount (N=103), and the sociopragmatic appropriacy of the speech act worded in most of such cases revealed the consideration of the SCV not referred to by either of the participants. In 54 situations, all three SCVs were mentioned, and these were mainly contributed to the interaction by both participants. Example 3 offers a case of the collaborative noticing of power, distance and imposition by Students A and B. While “position,” i.e. *power*, is referred to in the first turn by Student A, the other two aspects of the situation, i.e. *imposition* and *distance*, are noticed and verbalized in a subsequent turn by Student B.

**Example 3:**

**Apology situation:** You are a teacher. You promised one of your students to bring him/her a book on Wednesday afternoon, but you forgot. The student waited for you at the door of your office for one hour. Today is Thursday, and the student comes to your office again; you apologize to him/her.

**Paired interaction:**

- A. [reads the situation] Uhm, if I was the teacher, I would surely apologize, but not very formally or seriously, you know what I mean? it's because of the position, the teacher is the higher. Yeah?
- B. Yeah.
- A. So I wouldn't say that I'm sorry, I'm truly sorry, I forgot it. It's not ok to do so. What do you think?
- B. Yes, you're right. You know her position is much higher.
- A. Yes.
- B. You know, they have some power against us, the teacher. She doesn't have to talk about the reason; students should accept it. It seems the student is not very... very close, ... a normal relationship.
- A. I think for the teacher, the apology shouldn't be that formal. It is enough to say, "Sorry, I forgot to bring it." That's ok. Enough for me. As a student, I would accept it. It's acceptable for me as a student.

Think-aloud protocols and paired interactions were also compared in terms of instances of noticing ILP knowledge gaps, i.e. gaps in one's knowledge of appropriate speech act strategies and semantic formulae. Learners in ILG explicitly admitted their lack of knowledge, as evident in the underlined part of the think-aloud protocol in Example 4. In this case, the learner tried to fill in the gap based on internal (i.e. own) feedback. Although she noticed power inequality, internal feedback (the underlined part) failed her in her choice of the right semantic formula. Failure to make the appropriate choice was also the case with the other twelve instances of knowledge gap noticing detected in individual languaging.



**Example 4:**

**Request situation:** You've been working in an advertising company for five years. Having worked hard for the company's benefits and received better salary offers from other companies, you think you can ask the boss for a pay rise or promotion. You make this request.

**Think-aloud protocol:** In this situation, I want my boss to increase my salary and give me a better position, but he might refuse this request because his power is certainly more than mine. I should be very careful and formal, and make my request in a way that he does not feel I don't respect him. I can use "I...uhm" but I think my... I don't know what to say...how to be formal and at the same time effective in this situation.... Maybe, I should... I should say "I'm... I have worked very hard for this company"... "I expect a pay rise or a promotion, Sir!" If I say "Will you give me a pay rise?" it would not work. I think making this request is hard because the boss is more powerful.

In collaborative languaging, on the other hand, the instances of noticing ILP knowledge gaps were realized by implicit and explicit requests for information from partners. The explicit request for information has been underlined in the paired interaction in Example 5. In this example, the learner made the right choice based on her partner's feedback.

**Example 5:**

**Apology situation:** You are a university professor. Standing in the university hall, you are talking to one of your students about a project. In the meantime, one other student, who is very happy to see you after about a year and whom you really like, comes forward and pulls out his hand to shake hands with you. You just greet him, but do not notice his hand. He seems to have taken offence. You apologize.

- A. The first thing that I...I think is that I show in my face that I did not have any intention to upset him. And I think about how I can apologize. He is my old student. But I say, for example, "John, I'm so... so sorry, I was so busy with the other student that I failed to shake hands with you." I'm so sorry or I hope you forgive me? Which one is better or more appropriate?
- B. What?

- A. I'm so sorry or I hope you forgive me?
- B. The first is better because I think. The second one is too formal and the teacher does not have to be ashamed. It was not intentional if I'm not mistaken.
- A. So "I'm so sorry, I was so busy with the other student that I failed to shake hands with you." I agree ... This is better if we consider the teacher's power, and maybe what has happened is not very important.

Overall, more SCVs were noticed by the collaborative languaging group. In addition, ILP knowledge gaps were not noticed in most cases, irrespective of the type of languaging; however, the few observed instances remained unresolved in individual languaging, but resolved through external feedback in collaborative languaging.

### **Hypothesis Testing in Individual and Collaborative Languaging**

With respect to "hypothesis generation and testing," as one of the postulated learner output functions, there were 42 such episodes altogether: 12 in the individual and 30 in the collaborative languaging data. Hypothesis testing was defined as individual or collaborative trial-and-error postulations of various aspects of the situation (e.g., SCVs, SCVs' interactions, and SCV-politeness interaction) or appropriate speech act strategies and semantic formulae, induced by either internal or external feedback, leading to output modifications.

As for individual hypothesis testing episodes, learners drew on internal feedback in their output modifications. In Example 6, the learner modified her speech act strategies upon mulling over the consequence of performing the trialed speech act, as the underlined part of the think-aloud protocol shows. In fact, further reflection on the appropriacy and adequacy of her postulated speech act strategies seems to have pushed her to modify her output.

#### **Example 6:**

**Apology situation:** You borrowed a book from your classmate. While you were reading the book, you accidentally spilled some orange juice on the cover of the book. Now you return the book to your classmate and apologize.

**Think-aloud protocol:** In this situation [reads the situation], it's my fault ... really ... that I couldn't ... keep her book, and it maybe ... I should ... I should buy a new book for him. "Sorry I'll buy a new one for you," and I'm not sure that in this situation he

forgives me for... she forgives me if I say this, but I ... it's... it's her book, and I'm really sorry about what I did. I should say the decision is with her. I'll tell her I'll do everything that she decides, Yes. It's better. So I'll say "I'm really sorry about what I did. I don't know how it happened. Now, I'll do whatever you say, and I'm ready to buy a new book for you."

Regarding collaborative languaging, output modifications were induced solely by external feedback received from one's partner. In Example 7, Student A improved her first statement and received positive feedback from Student B.

**Example 7:**

**Request situation:** You are doing your research project, and need to interview the president of your university. The president was your teacher and you know him quite well. You know the president is very busy and has a very tight schedule. You still want to ask the president to spare one or two hours for your interview.

**Paired interaction:**

- A. We can say uhm "Would you please spare one or two hours for our interview, for my interview?"
- B. It's ... I think it is not enough, I mean the way you are requesting "would you please" or "could you please." It's ... it's the company president.
- A. Ok... then... we may say "I was wondering if you could spare one or two hours for our interview."
- B. Yeah, it's better...

In sum, collaborative hypothesis testing was more frequent, and was led by external feedback, rather than internal feedback.

### **Reflection in Individual and Collaborative Languaging**

An inspection of the data brought to light several instances of reflection (see Table 2). Regarding individual languaging, 78 reflections over the sociopragmatic aspects of the situation mainly prior to wording the speech acts, but also after that, were detected. Such reflections were characterized

by the learner contemplating whether or not to perform the speech act, possible consequences of its performance, situation-specific SCVs and their interaction, and politeness and its interaction with SCVs. It is worth noting that of the 15 participants, 3 failed to reflect over the issues of politeness and formality, while the other 12 made an explicit mention of them, at least in one of the 10 situations each dealt with. The underlined part of Example 8 offers a case of reflection over the situation and its associated SCVs.

**Example 8:**

**Apology situation:** It is the first session of a new course at university. Upon entering the class, you bump into one of your new classmates who is standing at the door talking on the phone. How would you apologize?

**Think-aloud protocol:** I come across this situation that... the person that I should apologize to is of the same level and in the same class of... in the same class of society with me, so it wouldn't be that hard to make this apology, so I would... but I think I should be polite as always, so I would say to him "I'm sorry! I wasn't looking." I think this is enough. Nothing important has happened.

Concerning collaborative languaging, the 80 reflection episodes detected were mainly distributed over several turns, with either of the participants having ideas to contribute about the sociopragmatic and pragmalinguistic aspects of the situation, not referred to by the other or pointed out in previous turns. In other words, "collaborative reflection" involved either initiating new ideas or providing greater depth to already postulated aspects of the situation by one's partner; this was where collaborative dialoging evolved. Comments followed by confirmation checks, i.e. checking whether or not the partner agreed, were common. In 5 cases, however, where interaction was minimal, reflections were almost totally spelled out by one of the partners, and the other either confirmed her ideas without any further comments or simply kept silent. Such cases were not counted as instances of "collaborative reflection." Example 9 is an excerpt of a paired interaction, illustrating collaborative reflection over the situation, politeness, and appropriate speech act strategies.

**Example 9:**

**Request situation:** You are doing your research project, and need to interview the president of your university. The president was your teacher and you know him quite well. You know the president is very busy and has a very tight schedule. You still want to ask the president to spare one or two hours for your interview.

**Paired interaction:**

- A. [Reads the situation]
- B. Oh, it is very difficult. You are so busy and we need two or three hours of his time.
- A. Aha!
- B. I think that we should be so polite.
- A. Yes...
- B. So thankful!
- A. Yes, and we should insist on our request. Why...
- B. We should or we shouldn't?
- A. Yes, we should but because we need it. But you know it is maybe... it is not polite.
- B. Yes, it is so difficult. I think that we should make our request for many times until he accepts.

Moreover, collaborative reflections led in some cases to the noticing of potentially important aspects of the situation not specified in the situation prompts, such as *gender* and *distance*. Another observation was “comparative reflection” incidents after discussing all the 5 speech act-specific situations. These occurred in 3 of the 15 paired interactions, one of which is presented in Example 10.

**Example 10:****Collaborative comparative reflection over request situations:**

- A. And I think that it depends on the situation how we should express our request: For example when we encounter with the president, we should be very polite, and we should manage our speaking, but when we want to speak, for example, with our roommate, it is not necessary to be very formal. Because our roommate is someone that he is... he or she is like us, the company's president is a very important person, or our teacher, our teacher...

- B. For example, the president is not that friendly with us, so we should be more polite probably, because it's an important person, and we don't have a friendly relationship.

The results of comparing individual and collaborative languaging in terms of their potential for enhancing noticing, reflection, and hypothesis testing can be summarized as follows:

1. Instances of noticing were frequent in both individual and collaborative languaging, though more SCVs were noticed for each noticing instance in the latter. With regard to noticing knowledge gaps, only paired participants managed to fill in the postulated gaps, rarity of such instances in both languaging types notwithstanding.
2. Hypothesis testing episodes were more frequent in collaborative languaging; they were induced by external feedback in collaborative languaging and by internal feedback in individual languaging. In addition, hypothesis testing in collaborative languaging was more clearly conducive to output improvements.
3. Collaborative reflections were more profound than individual reflections owing to the contribution of both participants; furthermore, comparative reflections characterized only collaborative languaging.

### **Discussion**

A comparison of think-aloud protocols of the individual languaging group (ILG) and paired interactions of the CLG in terms of Swain's (1985, 2006) postulated functions of learner output brought to light a number of differences. With regard to the noticing function of languaging, collaborative languaging led to the noticing of more SCVs for each situation prompt, compared with individual languaging. This observation was expected since "knowledge pooling," i.e. knowledge co-construction in a shared activity, induced by collaborative languaging is likely to lead to a more profound analysis of relevant aspects (pragmalinguistic or sociopragmatic) of the situation, implicating in speech act performance. From a sociocultural perspective, "the co-construction of linguistic knowledge in dialogue is language learning in progress" (Swain & Lapkin, 1998, p. 321). It follows that while ILG probably developed an awareness of such issues as power, distance, and imposition based solely on their own resources, CLG had the additional opportunity of "scaffolded help" of a peer, in the sense of bringing to light critical sociopragmatic

features, otherwise passed unnoticed (Mitchell & Myles, 2004, p. 197). Regarding “noticing the gap” (Schmidt & Frota; cited in Ellis, 2008), in the sense of recognizing a hole in one’s pragmalinguistic and sociopragmatic knowledge, gap filling or output improvements following peer feedback or comments featured in collaborative, but not in individual, languaging. This could be justified with reference to the nature of teacher-imposed languaging as a task type. Shehadeh (1999) found one-way tasks superior to two-way tasks in terms of the creation of output modification opportunities. In the present study, however, improved versions of earlier output featured in collaborative languaging. Collaborative languaging was, in effect, two-way in terms of “interactants’ relationship,” and convergent and collaborative in terms of “task orientation” (Ellis, 2003). Two-way tasks can be said to have fulfilled the potential for effecting improved verbalizations of earlier speech acts in the present study; however, how they compare with one-way tasks in terms of inducing speech act modification opportunities stands in need of research.

With respect to reflection episodes, collaborative reflections proved to be more profound, probably as a result of the “dialogically constituted interpsychological mechanism” (Donato; cited in Mitchell & Myles, 2004, p. 197). Collaborative reflections involved effective peer scaffolding, in the sense of initiating ideas regarding various social and linguistic aspects of speech act performance or building upon those already put forth by one’s partner. In line with Ohta (2000, 2001), the observed difference between individual and collaborative languaging can be discussed in terms of such cognitive concepts as “selective attention” and “L2 processing capacity” (Long, 1996, p. 414). CLG probably brought together such resources effectively when engaged in dialogic collaboration regarding the pragmalinguistically and sociopragmatically appropriate production of the speech act in question. This unique opportunity, however, was not available to ILG. Instead, they had to rely on their own limited working memory and processing capacity, hence their less-than-perfect reflections. In addition to collaborative reflections, CLG engaged, on occasion, in “comparative reflections,” though no such instances were observed for ILG. This finding can be explained in terms of Flavell’s (1979) concept of “metacognitive experiences,” defined as “any conscious cognitive or affective experiences that accompany or pertain to any intellectual enterprise” (p. 906). Comparative reflections featuring in paired interactions might be indicative of the participants’ metacognitive experiences: conscious comparisons of sociopragmatic features of the situations and of the ways their idiosyncratic functional and contextual features could be

mapped onto pragmalinguistic forms. This, in turn, probably shows the greater metacognitive awareness-raising potential of collaborative languaging.

Finally, think-aloud protocols and paired interactions were compared in terms of hypothesis generation and testing episodes. Such instances were not only more frequent in collaborative languaging, but also more clearly leading to output improvements. The main reason for this finding could be the presence of external (peer) feedback, distinguishing the two languaging types. In Vygotskian terms, such feedback can be thought of as a scaffold, which can function to sustain motivation and interest during problem-solving (Mitchell & Myles, 2004). The obtained results concerning the greater potential of collaborative languaging for learner reflection and hypothesis testing is in accordance with Storch's (2005, 2007) finding that collaborative output leads to a higher number of language-related episodes, compared with individual output. The results might also explain Khatib and Ahmadi Safa's (2012) finding as to the significant effect of co-equals' scaffolding on their speech act production and its superiority over teacher-fronted ZPD-wise scaffolding.

### **Conclusion and Implications**

As the main foci of the present study, individual and collaborative languaging were found to differ in terms of their potential for pushing EFL learners to (a) notice the three situational variables of power, distance, and imposition, as well as own pragmalinguistic and sociopragmatic knowledge gaps, (b) generate and test hypotheses regarding form–function–context mappings, and (c) reflect over various aspects of the situation implicating in speech act performance, including formality, politeness, and the interaction of these two factors with the three SCVs. While both individual and collaborative languaging can lead to noticing and reflection episodes, collaborative languaging tends to house a larger number of hypothesis testing episodes. In addition, collaborative noticing and collaborative and comparative reflection episodes tend to be more profound, owing to the availability of external (peer) feedback and to pragmatic knowledge pooling. Finally, collaborative hypothesis testing more clearly induces modification (improvement) of earlier output, and the noticed knowledge gaps can be better resolved in collaborative dialoging. According to Swain (2006), requiring learners to produce language in pairs or groups potentially yields collaborative



metalinguistic talk, in which they strive to thrive in the linguistic showcase. Individual production, on the other hand, does not come up to comparable standards.

Based on the findings of the present study, the socioculturalism-informed output hypothesis, which marries cognitive psychology and social practice theories, might have the potential to theoretically explain ILP development. This is despite the fact that in most related ILP studies, languaging has been referred to as a theoretical condition for speech act development alongside structured input, negative evidence, and meaning negotiation potentially engendered in interaction (Kasper, 2001; Martínez-Flor & Usó-Juan, 2010). That mere languaging, in the absence of explicit or implicit instruction, can aid learners in their endeavor to learn L2 pragmatic features has been evidenced in few studies (Khatib & Ahmadi Safa, 2012; Tajeddin & Bagherkazemi, 2014). These studies are generally in favor of collaborative dialoguing, and the present study's results partially explain the reason for this tendency; however, more studies addressing the nature of languaging from various angles including the significance of the nature of grouping (proficiency grouping; ILP expert peers or ILP co-equals; members' expressiveness and willingness to communicate), the language and modality of languaging (L1 or L2; oral or written), and learners' reference to mediating artifacts (dictionaries, the Net, etc.) are needed to draw a generalizable conclusion. Overall, it seems to be high time ILP practitioners disengaged themselves from the haunting dilemma of implicit or explicit pragmatic instruction and propelled their practices into a consideration for learners' own potential, including their individual and collaborative languaging.

## References

- Bagherkazemi, M. (2014). *Interlanguage Pragmatic Development: Impacts of Individual Output, Collaborative Output, Input Enhancement, Metapragmatic Awareness Raising, and Pragmatic Learning Strategies* (Unpublished doctoral dissertation). Allameh Tabataba'i University, Tehran, Iran.
- Bao, G. (2019). Input and output tasks in EFL learners' vocabulary acquisition. *TESOL International Journal*, 14(1).
- Bardovi-Harlig, K. (2001). Evaluating the empirical evidence: Grounds for instruction in pragmatics? In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 13-32). Cambridge: Cambridge University Press.
- Brooks, L., Swain, M., Lapkin, S., & Knouzi, I. (2010). Mediating between scientific and spontaneous concepts through languaging. *Language Awareness*, 19(2), 89-110.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching. *Applied Linguistics*, 1(1), 1-47. doi: 10.1093/applin/I.1.1
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford: Oxford University Press.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive development enquiry. *American Psychologist*, 34(10), 906-911. doi:10.1037/0003-066X.34.10.906
- Geranpayeh, A. (2003). A quick review of the English Quick Placement Test. *University of Cambridge ESOL Examinations Research Notes*, 12, 8-10. Retrieved from [http://www.uniss.it/documenti/lingue/what\\_is\\_the\\_QPT.pdf](http://www.uniss.it/documenti/lingue/what_is_the_QPT.pdf)
- Ishikawa, M. (2015). Metanotes (written languaging) in a translation task: do L2 proficiency and task outcome matter? *Innovation in Language Learning and Teaching*, 9(2), 115-129.
- Ishikawa, M., & Suzuki, W. (2016). The effect of written languaging on learning the hypothetical conditional in English. *System*, 58, 97-111.
- Jernigan, J. E. (2007). *Instruction and developing second language pragmatic competence: An investigation into the efficacy of output* (Doctoral dissertation). Available from ProQuest

Dissertations and Theses database. (UMI No. 3263885)

- Jia, D. (2015). *An Experimental Study on the Effect of Written Languageing on English Writing Revision* (Unpublished master's thesis). Lanzhou University, Lanzhou.
- Kasper, G. (2001). Classroom research on interlanguage pragmatics. In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 33–60). Cambridge: Cambridge University Press.
- Khatib, M., & Ahmadi Safa, M. (2011). The effectiveness of ZPD-wise explicit/implicit expert peers and co-equals' scaffolding in ILP development. *Iranian Journal of Applied Linguistics (IJAL)*, 14(1), 49-75.
- Knouzi, I., Swain, M., Lapkin, S., & Brooks, L. (2010). Self-scaffolding mediated by languaging: Microgenetic analysis of high and low performers. *International Journal of Applied Linguistics*, 20(1), 23-49.
- Krashen, S. (1985). *The input hypothesis: Issues and implications*. Harlow, UK: Longman.
- Lantolf, J. (2011). The sociocultural approach to second language acquisition: Sociocultural Theory, second language acquisition, and artificial L2 development. In D. Atkinson (Ed.), *Alternative approaches to second language acquisition* (pp. 24-47). London / New York: Routledge.
- Li, L. (2015). Effect of languaging on the acquisition of tense and aspect. *Papers on Education*, 19, 22-25.
- Liang, Z. (2014). *Effect of Written Languageing on Second Language Writing* (Unpublished master's thesis). Xiangtan University, Xiangtan.
- 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). Waltham, MA: Academic Press.
- Martinez-Flor, A., & Usó-Juan, E. (2010). Pragmatics and speech act performance. In A. Martinez-Flor & E. Usó-Juan (Eds.), *Speech act performance: Theoretical, empirical and methodological issues* (pp. 3-20). Amsterdam: John Benjamins.
- Mitchell, R., & Myles, F. (2004). *Second language learning theories* (2nd ed.). London: Hodder Arnold.
- Moradian, M. R., Miri, M., & Hossein Nasab, M. (2017). Contribution of written languaging to enhancing the efficiency of written corrective feedback. *International Journal of Applied*

- Linguistics*, 27(2), 406-421. doi: 10.1111/ijal.12138
- Negueruela, E., & Lantolf, J. (2006). Concept-based pedagogy and the acquisition of L2 Spanish. In R. Salaberry & B.A. Lafford (Eds.), *The art of teaching Spanish: Second language acquisition from research to praxis* (pp. 79–102). Georgetown: Georgetown University Press.
- Niu, R. (2017). The Mediation of EFL learners' interaction for word processing within collaborative output. *TESOL International Journal*, 12 (2).
- Niu, R., & Li, L. (2017). A review of studies on languaging and second language learning (2006-2017). *Theory and Practice in Language Studies*, 7(12), 1222-1228. DOI: <http://dx.doi.org/10.17507/tpls.0712.08>.
- Norouzian, R., & Eslami, Z. (2016). Critical perspectives on interlanguage pragmatic development: An agenda for research. *Issues in Applied Linguistics*, 20(1), 25-50.
- Ohta, A. S. (2000). Rethinking interaction in SLA: Developmentally appropriate assistance in the zone of proximal development and the acquisition of L2 grammar. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 51–78). Oxford: Oxford University Press.
- Ohta, A. S. (2001). *Second language acquisition processes in the classroom: Learning Japanese*. Mahwah, NJ: Lawrence Erlbaum.
- Ohta, A. S. (2005). Interlanguage pragmatics in the zone of proximal development. *System*, 33(3), 503–517. DOI: 10.1016/j.system.2005.06.001
- Rose., K. R., & Kasper, G. (Eds.), *Pragmatics in language teaching*. Cambridge: Cambridge University Press.
- Schmidt, R. (1993). Consciousness, learning, and interlanguage pragmatics. In G. Kasper & S. Blum-Kulka (Eds.), *Interlanguage pragmatics* (pp. 21–42). Oxford: Oxford University Press.
- Shehadeh, A. (1999). Non-native speakers' production of modified comprehensible output and second language learning. *Language Learning*, 49(4), 627–675. doi: 10.1111/0023-8333.00104
- Shehadeh, A. (2002). Comprehensible output, from occurrence to acquisition: An agenda for acquisitional research. *Language Learning*, 52(3), 597–647. doi:10.1111/1467-9922.00196
- Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing*, 14(3), 153–173. doi: 10.1016/j.jslw.2005.05.002
- Storch, N. (2007). Investigating the merits of pair work on a text editing task in ESL classes.

- Language Teaching Research*, 11(2), 143–59. doi: 1177/1362168807074600
- Suzuki, W. (2009). Improving Japanese university students' second language writing accuracy: Effects of languaging. *Annual Review of English Language Education in Japan*, 20, 81-90.
- Suzuki, W. (2012). Written languaging, direct correction, and second language writing revision. *Language Learning*, 62(4), 1110-1113.
- Suzuki, W. (2017). The effect of quality of written languaging on second language learning. *Writing & Pedagogy*, 8(3), 461-482.
- Suzuki, W., & Itagaki, N. (2009). Languaging in grammar exercises by Japanese EFL learners of differing proficiency. *System*, 37, 217-225.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Rowley, MA: Newbury House.
- Swain, M. (2006). Languaging, agency and collaboration in advanced second language proficiency. In H. Byrnes (Ed.), *Advanced Language Learning: The contribution of Halliday and Vygotsky* (pp. 95–108). London: Continuum.
- Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal*, 82(3), 320–337. DOI: 10.1111/j.1540-4781.1998.tb01209.x
- Swain, M., & Lapkin, S. (2001). Focus on form through collaborative dialogue: Exploring task effects. In M. Bygate, P. Skehan & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 99–118). Upper Saddle River, NJ: Pearson Education.
- Taguchi, N. (2006). Analysis of appropriateness in a speech act of request in L2 Spanish. *Pragmatics*, 16(4), 513–533.
- Taguchi, N. (2011). Teaching pragmatics: Trends and issues. *Annual Review of Applied Linguistics*, 31, 289–310. doi: 10.1017/S0267190511000018
- Taguchi, N. (2015). Instructed pragmatics at a glance: Where instructional studies were, are and should be going? *Language Teaching*, 48, 1-50. DOI: 10.1017/S0261444814000263
- Tajeddin, Z., & Bagherkazemi, M. (2014). Short-term and long-term impacts of individual and collaborative pragmatic output on speech act production, *Teaching English Language*,

- 8(1), 141-166.
- Tajeddin, Z. & Tayebipour, F. (2012). The effect of dynamic assessment on EFL learners' acquisition of request and apology, *Journal of Teaching Language Skills*, 31(2), 87-118. DOI: <http://dx.doi.org/10.22099/jtls.2012.499>
- Takimoto, M. (2009). The effects of input-based tasks on the development of learners' pragmatic proficiency. *Applied Linguistics*, 30, 1–25.
- Takimoto, M. (2010). Evaluating the effects of task repetition on learners' recognition and production of second language pragmatic chunks. Paper presented at the International Conference of Malaysian English Language Teaching Association, Selangor, Malaysia.
- VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Norwood, NJ: Ablex.
- Zand Moghadam, A. (2012). *Pre-task, Post-task, and Acaffolded While-task Pragmatic Instruction: Effects on EFL Learners' Pragmatic Development* (Unpublished doctoral dissertation). Allameh Tabataba'i University, Tehran, Iran.