TESOL International Journal

Teaching English to Speakers of Other Languages

TESOL INTERNATIONAL JOURNAL

Volume 9 Issue 2, 2014

Contents

Foreword from the editor

Xinghua (Kevin) Liu

I am happy to open the new edition of our journal publication, which has five papers. In the first paper, **Hui Li** conducted a well-controlled experiment in which candidates' vocabulary production in oral proficiency interviews and its relation to raters' judgements of vocabulary were examined. The study presented many revealing findings, including the proficiency effect on candidates' spoken vocabulary and the relationship between the candidates' and the interviewer's lexical measures and the vocabulary scores assigned. This study has important theoretical and practical implications for oral production assessment. In the second paper, **Analiza Liezl Perez-Amurao** studied the role of feedback and conferencing in the process writing approach by university students in the Philippines and Thailand. She found both groups of students showed a preference for the process approach and this approach had a positive effect upon the quality of students' final essays. **Jason Moser** presented in the third paper a very interesting study on the influence of transcribing, reporting, and task repetition on in-class student oral task performances. Through a well-designed experiment, the study did not find any significant difference between transcribing or reporting on subsequent task performances, but it suggested a significant task repetition effect on task performances. In the fourth paper, **Wilkinson Daniel O. Wong Gonzales** compared the effectiveness of two instructional approaches, namely the nonconventional learner-centered (NLC) and conventional teacher-centered (CTC) approaches in teaching Afro-Asian literature in Philippine private school context. The study found that the teacher preferred the CTC approach while students preferred the NLC approach though there were no significant differences in the Afro-Asian literature test scores immediately after the CTC or NLC instructional sessions. In the last paper, **Asmaa AlSaqqaf, Siti Jamilah Bidin** and **Ahmad Affendi Shabdin** investigated English language anxiety among Arab postgraduates studying at a higher education institution in Malaysia. The study revealed these sampled Arab postgraduates expressed differing levels of anxiety within the academic contexts and outside in everyday communication situations. They also discussed some interesting findings regarding the relationship between demographic variables and students' level of language anxiety.

At this time of the year, I would like to say thank you to our Editorial Board members for their support and hard work. Particularly I am thankful to the Team of Associate Editors (Andrew Pollard, Fan-Wei Kung, Hanh thi Nguyen, Khadijeh Jafari, Reza Dashtestani, Yu-Chih (Max) Lo) who deal with a large number of incoming submissions with great care and professionalism. Without their devotion, the journal can not stand as it is.

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

Candidates' Vocabulary Production and the Effect of Interviewer Accommodation on Vocabulary Scores in Oral Interviews

Hui Li
(Hui.Li@xjtlu.edu.cn)
Xi'an Jiaotong – Liverpool University, China

Abstract

This experiment examines candidates' vocabulary production in oral proficiency interviews (OPIs) and its relation to raters' judgements of vocabulary under such dialogic testing conditions. Twenty Chinese candidates' interviews were analysed in terms of lexical output (types and tokens), lexical diversity (D), and lexical sophistication (P_Lex and VocabProfile) vis-à-vis those of their interviewer's vocabulary output. Significant correlations were found between some of the candidates' and the interviewer's lexical measures and vocabulary scores assigned. The result also suggests a relationship between the interviewer's vocabulary production and candidates' vocabulary/ proficiency levels, implying that interviewer's support factored in candidates' vocabulary scores. The results also reveal strengths and weaknesses of various lexical measures in quantifying candidates' oral production.

Key words: Vocabulary production, assessment, lexical statistics, oral interviews, raters, judgment

Introduction

What we know of how vocabulary is assessed in oral examinations is not only scarce but also fragmented. Over the last two decades, whereas mainstream vocabulary studies have mainly contributed to research on written vocabulary, language testers have been more interested in testing communicative or interactional competence. Understanding on vocabulary assessment in oral examinations by and large was interpreted from a series of studies on related subjects, such as Brown, Iwashita, and McNamara's (2005, p. 2018) investigation of raters' perception on oral proficiency, Lim and Galaczi's (2010) examination of candiates' vocabulary at differing proficiency levels on Cambridge ESOL's Main Suite Examinations, and Lorenzo-Dus and Meara's (2005) examination of the relationship between candidates' vocabulary production and interviewer's linguistic accommodation in an oral proficiency interview. Jin and Mak's (2013) investigation of distinguised features in Chinese students' speaking performance. Vocabulary, in all the above studies, was found to be an effective indicator of oral proficiency.

Positive correlation found between vocabulary and language proficiency has stimulated an increasing interest among language testers to incorporate vocabulary as a criterion to assess oral proficiency. Only limited amount of research, however, has explored the relationship between candidates' lexical statistics and raters' judgements of vocabulary (Brown et al, 2005; Iwashita et al, 2008; Read, 2005). Although there are reports on positive correlations between candidates' vocabulary and raters' scores, these correlations are typically obtained by using holistic scores rather than using discrete vocabulary scores (c.f. Engber, 1995). In fact, when using discrete vocabulary scores, some studies (e.g. Malvern and Richards, 2002) have found insignificant correlations between candidates' vocabulary and raters' judgements on vocabulary. Investigation of the relation between candidates' lexical statistics and raters' scores is therefore necessary in order to provide empirical evidence of the reliability of vocabulary scores in oral examinations.

In response to the need of furthering the understanding of how vocabulary is assessed in oral examinations, this paper examines candidates' vocabulary performance in OPIs (oral proficiency interviews) in relation to their

overall language proficiency levels and interviewers' vocabulary production. It is hoped that the results could benefit both language testers and teachers, and shed insight into the interpretation of vocabulary scores in oral interview contexts.

Candidates' Performance in Oral Proficiency Interviews

Candidates' performance in OPIs is regarded as a collaborative achievement between interviewers and candidates. Quite a few studies have explored inter-interviewer variability. However, with the exception of Lorenzo-Dus and Meara's (2005) study, few studies have directly investigated the relationship between examiners' variation and candidates' vocabulary production. Consequently, it is not clear how interviewer variation (i.e. interviewer accommodation) affects candidates' lexical performance and vocabulary scores. Accommodation refers to the way interviewers 'modify the form and content of their discourse in order to facilitate communication' (Ross and Berwick, 1992, p. 162). Past studies have warned of the influence of interviewer accommodation on the interviewers'/raters' perceptions of candidates' performance (Ross and Berwick, 1992; Lazaraton, 1996). However, as Nakatsuhara (2008) states, little is known about how this threat is translated into scores on analytic rating scales. In her studies of the impact of inter-interviewer variability, she found that candidates' pronunciation and fluency were affected by interviewer differences. Vocabulary was among the five categories she examined. Although no significant difference was found between vocabulary scores in the two interviews conducted by different interviewers, there was a tendency for raters to mark this category lower when the interviewer style was 'teacherly' as opposed to 'formal'. Efforts, therefore, should be made to ascertain the kind of influence that the examiner-candidate interaction has on the latter's lexical performance and corresponding vocabulary scores. This in turn, can lead to a more reliable interpretation of candidates' vocabulary scores.

When speaking of assessing vocabulary, it should be noted that the construct of vocabulary is not clearly conceptualised in OPIs. The theoretical definitions of constructs in oral examinations, as Bachman and Palmer (1996, p. 212) state, either build on the 'content of a language learning syllabus or a theoretical model of language ability'. However, as Meara (1996, p. 37) writes about vocabulary assessment in general, 'the basic problem seems to be that we do not have a properly worked out theory of what constitutes lexical competence'. Vocabulary knowledge is multidimensional (Richards, 1976). Knowing a word involves knowing its form, meaning and use, both productively and receptively (Nation, 2002). When it comes to assessing spoken vocabulary, there seem to be dilemmas regarding what to test: size and/or quality (breadth and/or depth); knowledge and/or use.

The breadth versus depth debate has lasted over two decades and is still on-going (Meara, 1999; Qian, 1999; Vermeer, 2001). Measuring size, as argued by some researchers (Schmitt and Meara, 2000; Wesche and Paribakht, 1996), has its limitations on the ability to measure the extent to which a given word is known. Several studies on oral lexical proficiency, nevertheless, have largely focused on the size dimension of vocabulary, particularly on range of vocabulary/lexical richness (Daller and Xue, 2007; Duran, Malvern, Richards, and Chipere, 2004; Jarvis, 2002; Tidball and Treffers-Daller, 2008; van Hout and Vermeer, 2007; Vermeer, 2004; Yu, 2009). Likewise, range of vocabulary has recently been regarded by language testers as an important criterion in oral proficiency examinations. This overlap regarding the salience of lexical richness between research on vocabulary and in language testing is to be expected, in part at least due to the increasing cross-disciplinary collaboration, as shown in Read's (2005) involvement in IELTS revision and investigations of the test. Besides, range of vocabulary is conceived as (1) one of the salient factors that determine the quality of writing, and (2) a representation of the relationship between vocabulary knowledge and vocabulary use (Laufer and Nation, 1995). Since evidence is scarce with respect to the relationship between range of vocabulary and L2 oral proficiency, it is reasonable to suppose that the assumed relation of the two is based upon the understanding of them in written data. To take such a stance seriously, however, supporting empirical evidence is needed.

Regarding candidates' vocabulary production in OPIs, past studies have yielded mixed results. For example, some studies have found that lexical diversity is correlated with candidates' proficiency levels (Malvern and Richards, 2002); others have suggested, however, that D(measure of lexical diversity) does not discriminate between candidates' proficiency levels (Lorenzo-Dus and Meara, 2005). Lexical diversity refers to 'a variety of different words rather than a limited number of words used repetitively' (Read, 2000, p. 200). It is traditionally counted by types or Type Token Ratio (TTR). The latter has been the most popular, as well as the most criticized approach to calculating lexical diversity. TTR is the ratio of the number of different words (types) and total number of running words (tokens) used in a text. There are two main criticisms of this measure. First, it does not generate a constant value (Malvern and Richards, 1997; van Hout and Vermeer, 2007; Vermeer, 2000, 2004): there is no linear relationship between types and tokens. As language proficiency develops, types and tokens are found to develop at different rates. Tokens are known to increase relatively faster than types. Second, type/token ratio depends on the sample size: the longer the text is, the fewer new words will be produced (Malvern and Richards, 1997, 2002). So, shorter texts have higher TTR values than longer texts. TTR, therefore, may not produce reliable comparisons between language samples of various lengths, such as spontaneous oral production (Vermeer,2000; Daller and Xue, 2007).

In 1997, Malvern and Richards proposed a new approach to measuring lexical diversity using curve-fitting procedures. They used mathematical modelling to plot the progress of type/token ratio as tokens increase. Then the best fit between the theoretical and empirical curves was found by adjusting a parameter, D (ranging from 0 to 90). They claimed that D was not an index of text length, thus a reliable and valid measure of lexical diversity. This measure evoked great interest among researchers in the field of vocabulary study. It is nowadays a popular research tool for measuring lexical diversity (Duran, et al., 2004; Jarvis, 2002; Malvern, Richards, Chipere, and Durán, 2004; Silverman and Ratner, 2002; Yu, 2009).

A number of vocabulary measures, such as D mentioned above, have been developed in the last two decades. However, as Nation (2007) observes, little attention has been paid to how effective these measures are in written and/or oral contexts. He calls for studies to apply those measures to examine vocabulary in different contexts since they 'need to be checked by some other form of delivery' (Nation, 2007, p. 43). The oral context is one in particular need of close attention. Daller and Xue (2007) sought to investigate which lexical measure was most suitable for measuring oral proficiency of Chinese EFL learners. The measures examined in their study were divided into two categories: list-based measures of lexical sophistication (measures based on word frequency: LFP/Beyond 2000, PLex, and Advanced Guiraud) and list-free measures of lexical diversity (measures that were not based on word frequency: TTR, Guiraud Index, and D). By applying the two categories of vocabulary measures to Chinese EFL learners, the authors claimed that list-free measures were more effective than list-based measures in quantifying oral vocabulary.

Simply counting types and tokens, apparently, is not sufficient to display the difference in vocabulary. As such, judgement based solely on this type of calculation may produce biased interpretation of L2 learners' vocabulary. As van Hout and Vermeer state, 'both increasing length of sentences and variance in word frequency make any measure based on a relationship between types and tokens in a growing vocabulary very complicated' (van Hout and Vermeer, 2007, p. 100). The underlying assumption of list-based measures (Lexical Frequency Profile, Laufer and Nation, 1995; Guiraud Advanced, Daller, van Hout and Treffers-Daller, 2003; Measure of Lexical Richness (MLR), Vermeer, 2004; P_Lex, Meara and Bell, 2001) is that there is a link between vocabulary size and word frequency: the bigger vocabulary size is, the more low frequency words are used. Empirical evidence from past studies endorsed the assumption. For example, Richards and Malvern (2007) tested X_Lex (Meara and Milton, 2003) on twelve year-12 students. X_Lex was a list-based Yes/No vocabulary test designed to test low-level intermediate students. The students' mean words were shown to decline steadily from the highest frequency band to the lowest frequency band, which implied that word frequency is a sensitive factor in L2 acquisition.

Among all measures that are based on word frequency, two measures deserve our special attention: LFP

(Laufer and Nation, 1995) and P_Lex (Meara and Bell, 2001). LFP was the first attempt to use word frequency to test vocabulary development. Also, significant correlations between LFP and candidates' vocabulary growth have been recurrently reported (Laufer and Paribakht, 1998; Morris and Cobb, 2004; Muncie, 2002). However, disagreement on the effectiveness of the measure has also been noticed (Meara, 2005; Meara and Bell, 2001). In particular, Meara's Monte Carlo simulation of LFP raises a few questions about the reliability of LFP (see Meara 2005 for a detailed discussion). Meara (2005) confirmed that LFP worked well with candidates of distinct vocabulary size and of lower proficiency levels, but cautioned about its possible sensitivity and stability in gauging modest vocabulary change.

P_Lex (Meara and Bell, 2001) is designed to (1) estimate 'difficult' words used (beyond the first 1,000 content words); (2) yield one single figure of lambda (value between 0 and 4.5); and (3) work with short texts (longer than 120 words). P_Lex is produced on the principle known as 'Zipf's law' in that 'frequency multiplied by rank has a constant value' (Vermeer, 2000, p. 78). Mixed findings emerged in the literature regarding how P_Lex discriminated between candidates by proficiency levels. Some studies found significant results (Bell, 2003; Miralpeix, 2007); while others did not (Espinosa, 2005). In particular, Espinosa (2005) examined the effect of text length on PLex by comparing and correlating it between (a) a fixed number of tokens and (b) full text length of various numbers of tokens. The correlation between the two computations were significant (r = .828, p<.0001), but a significant mean difference between them was also found (t=3.88, p<.0001). This finding confirmed that P_Lex was effective with short text length, but also revealed that P_Lex was not completely independent of text length: longer texts have higher lambda values than shorter texts.

None of the measurement tools examined above provides a 'fit for all' solution to how best to investigate candidates' spoken vocabulary production. Each measure has its own advantages and disadvantages. Different measures are not competitive. Rather, they are complementary, tapping into distinct aspects of vocabulary knowledge. Thus, 'having several measures provides a more comprehensive and thus useful picture of vocabulary knowledge' (Nation, 2007, p. 42).

In this paper, various lexical measurements are used to investigate candidates' lexical production in OPIs vis-à-vis an interviewer's vocabulary output. In a nutshell, the experiment addresses the following research questions:
(1) Does proficiency affect vocabulary output?
(2) Is vocabulary score a function of interviewer-candidate interactions in the OPIs?

Methodology

Candidates
The data included 20 OPIs taken by Chinese students at a university in the UK. Each interview lasted between five to ten minutes, giving a total of 15,154 words of 20 transcripts. Among the 20 candidates, ten candidates (group A) were at the time of the experiment enrolled in Master courses in translation and media studies; while another 10 candidates (group B) were taking EFL classes prior to starting degree courses at the university. The candidates' IELTS scores, which were taken four months prior to the experiment, were used as indicators of the candidates' language proficiency. Independent samples t-test shows that the IELTS scores (group A: mean = 6.35; group B: mean = 4.65) in the two groups are statistically significant (t = 10.100, p<.001, one-tailed).

Interviewer
One interviewer conducted all interviews for both of the two proficiency groups. The interviewer was also the teacher of all candidates. With more than ten years' experience in teaching EFL courses, she was also a trained oral examiner (a certified IELTS examiner). In this experiment, she had the double role of interviewer and rater. After the tests, another IELTS certified English native speaking rater double marked the whole data set. The inter-rater reliability was computed using Spearman rank order correlation. The correlation for the entire OPI scores was statistically significant (rho= .834, p<.001).

As previously known, different interviewers vary in their interview styles (Brown, 2003; Lazaraton, 2002; Ross, 1992, 2007), which may result in different amount of support to candidates. One strength of this experiment, therefore, is that candidates were subject to a single interview style rather than various interview styles that would have been met if multiple interviewers had been used.

Materials
The questions used in the OPIs were adopted from the OPI format in BAF (Barcelona Age Factor) Project of Barcelona University (1995 – 2002). The main purpose of that project was to examine the effect of age on L2 language learning (Muñoz, 2006). Altogether, the OPIs included 15 questions. The interviewer was not prevented from intervention during OPIs. Past research (e.g. Ross, 1988) has shown that strict framework may influence test-takers' performance and may even be 'detrimental to candidates'. The examiner prepared a question plan and was told to facilitate natural conversation with test-takers.

Construct
Following Malvern and Richards' (2002) study, six constructs were assessed in the OPIs in this experiment, namely, range of vocabulary, fluency, complexity of structure, content, accuracy and pronunciation. The examiner defined the constructs prior to the experiment (see Table 1) and rated them accordingly. The rating was based on a six-point scale (1 poor and 6 completely satisfactory). The overall OPI scores were calculated as the total score of the six constructs, and were used as an index of candidates' oral proficiency.

Table 1
The Examiner's Understanding of the Six Measures Used in the Test

Range of vocabulary	The variety of vocabulary use, such as common use vocabulary and more specific vocabulary.
Complexity of structure	The level of difficulty present within a sentence or grammatical structure.
Fluency	The quality of speech delivery; the level of smooth speech.
Content	The level of difficulty contained within the sentences being constructed in the interview (including cohesion).
Accuracy	Tests the level of errors identified within the speech.
Pronunciation	The level of accuracy with which English language constructions and sounds are pronounced.

Test Administration
The interviews were conducted in English and lasted about eight minutes on average. They were conducted as a one-to-one interaction. All OPIs were digitally recorded, with the prior agreement of the candidates and the examiner. The examiner was also asked to comment on candidates' vocabulary after each interview. These comments were digitally recorded and used as supplementary data for analysis.

Data Analysis
The 20 digital interviews were transcribed and coded in accordance with the conventions in Codes for the Human Analysis of Transcripts (CHAT) of the Child Language Data Exchange System (CHILDES) project (MacWhinney, 2000). The utterances were segmented on the basis of semantic-based T-units, which are defined by Young and Milanovic (1992, p. 409) as 'an independent clause and any dependent clauses.' Elements to be

counted as one T-unit include: a single clause, a main clause plus subordinate clauses, one fragment produced by ellipsis, and two or more phrases in apposition (Young, 1995).

All transcripts were run through the check command of the Computerized Language Analysis (CLAN) program to ensure they were viable for further analysis. Following past studies (Brown, Iwashita, and McNamara, 2005; Malvern and Richards, 2002; Read, 2005), non-completed words, proper nouns, immediate self-repetitions, discourse continuer (*ok, mm* etc.) and paralinguistic codes were excluded from the subsequent analysis.

Measures of lexical size (such as lexical diversity and lexical sophistication) were calculated for each of the 20 transcripts both for the interviewer's and the candidate's discourses. The following lexical measures were calculated: types, tokens, TTR, D (Clan), P_Lex (P_Lex1.1) and Vocabulary profile (Cobb, 2007). Components of vocabulary profile includes vocabulary frequency for the first 1,000 words (K1), the second 1,000 words (K2), and off-lists (including AWL and words not in the first 2K).

D was calculated twice by using two programs: the vocd command of the CLAN program (MacWhinney, 2000) and D-tools (Meara and Miralpeix, 2004) to ensure the reliable computation of D statistics. A statistically high correlation was found (candidates' Ds: rho = .977; p< .01; examiner's Ds: rho = .867; p< .01) between the two sets of D calculated by the two programs. This evidence provided support for the reliability of D values obtained from CLAN, and a decision was made to use it for statistical analysis in this experiment. Note that inflectional D was used. That means, for example, that fly, flies, and flying were regarded as different types.

Results

RQ1: Does proficiency affect output?
All the lexical measures examined were compared between the two groups (Table 2) on an Independent samples t-test (Table 3). Significant differences were found for some of the measures, namely, types (t = 3.275, p<.01), tokens (t = 3.288, p<.01), P_Lex (t = 4.037, p<.001) and off-lists (t = 2.785, p<.05), with P_Lex being the most significant. TTR, D, K1, and K2 were not significantly different between the two proficiency groups.

Table 2
Descriptive Statistics for Candidates' Vocabulary Measures

	Group A		Group B	
	Mean	Std.	Mean	Std.
Types	161.40	66.08	89.20	22.24
Tokens	377.10	189.27	171.10	58.57
TTR	0.46	0.09	0.53	0.06
D	67.88	12.99	60.05	5.56
PLex	0.68	0.22	0.34	0.16
K1	90.93%	0.96%	92.35%	2.28%
K2	3.91%	1.25%	3.82%	1.91%
Off-lists	5.15%	0.97%	3.83%	1.14%
IELTS	6.35	0.41	4.65	0.34

Table 3
Independent Samples T-test for Candidates' Vocabulary Measures (Group A –Group B)

						95% Confidence Interval of the Difference	
						t-test for Equality of Means	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Types	3.275	18	.004	72.20	22.05	25.879	118.521
Tokens	3.288	18	.004	206.00	62.65	74.369	337.631
TTR	-2.029	18	.058	-.07	.03	-.142	.002
D	1.753	18	.097	7.83	4.47	-1.557	17.223
Plex	4.037	18	.001	.34	.08	.165	.521
K1	-1.810	18	.087	-1.42	.78	-3.06	.23
K2	.125	18	.902	.09	.72	-1.43	1.61
Off-lists	2.785	18	.012	1.32	.47	.32	2.32

Candidates' vocabulary frequency profiles are further illustrated in Figure 1. As seen in the figure, group A candidates produced fewer words at K1 level than group B candidates did, but the difference was not significant (Table 2 & 3).   Also, no difference was found between groups at the K2 level. Off-lists words, which include academic vocabulary list (AWL) and other words, only constituted 5.16% of the total in group A, and 3.83% in group B respectively – the difference being moderately significant. Candidates in both groups used a considerable number of high frequency words (K1 and K2).
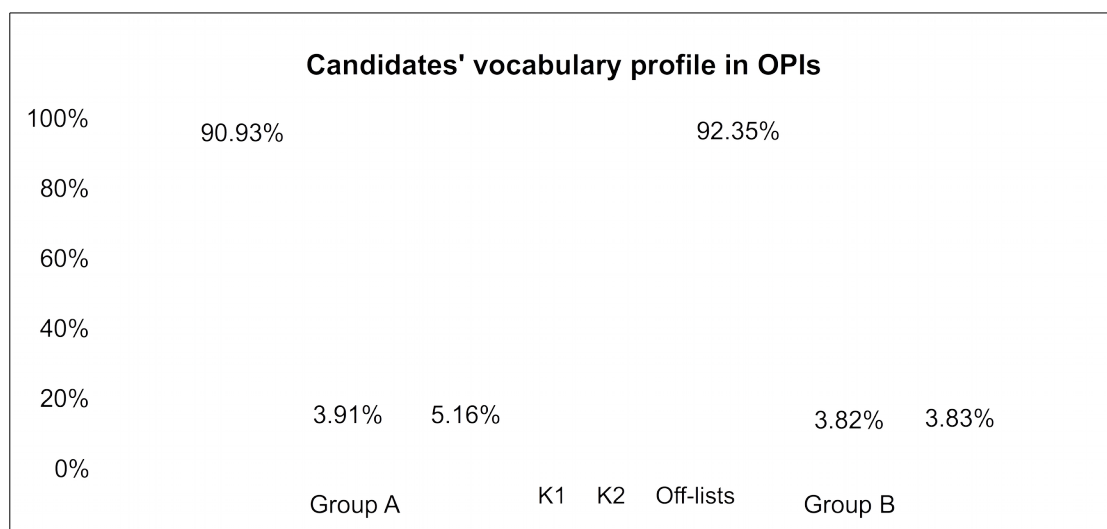

Figure 1. Vocabulary frequency profile for candidates in group A and B

Spearman rank order correlation analysis was run between different lexical measures and vocabulary scores, oral proficiency scores (OPI scores) and overall language proficiency scores (IELTS score). As shown in Table 4, apart from K1 and K2 words, all other measures correlated significantly with vocabulary scores, OPI scores and IELTS scores. Note that the correlation between candidates' lexical statistics and vocabulary scores was only moderate and was not stronger than the former with the other two scores (OPI scores and overall language proficiency scores).

Table 4
Correlations Between Candidates' Vocabulary Measures and Their Proficiency Measures (n=20)

		Types	Tokens	TTR	D	Plex	K1	K2	Off-lists
Vocabulary	rho	.593**	.575**	-.410	.481*	.693**	-.217	-.242	.673**
	Sig. (2-tailed)	.006	.008	.073	.032	.001	.358	.304	.001
OPI	rho	.585**	.605**	-.513*	.511*	.721**	-.173	-.129	.489*
	Sig. (2-tailed)	.007	.005	.021	.021	.000	.467	.587	.029
IELTS	rho	.567**	.563**	-.476*	.473*	.710**	-.249	-.069	.524*
	Sig. (2-tailed)	.009	.010	.034	.035	.001	.290	.772	.018

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4 also shows that P_Lex had the highest correlation with vocabulary scores (rho = .693, p<.001), OPI scores (rho=.721, p<.001) and IELTS scores (rho= .710, p<.001), indicating that P_Lex was a salient lexical measure in vocabulary production and assessment in the OPIs. However, it is interesting that PLex, being a measure of lexical sophistication, correlated more strongly with OPI scores and IELTS scores than with vocabulary scores.

RQ2: Is vocabulary score a function of interviewer-candidate interactions in the OPIs?

Means of the interviewer's and candidates' lexical measures are displayed in Table 5. There was an overall decline of values from candidates to interviewer in each of the two groups. This means that the interviewer produced fewer words, fewer different words, fewer rare words, and a smaller range of vocabulary variation than the candidates, especially when she paired with group A candidates.

The interviewer's lexical measures were further entered into independent-samples t-tests (two-tailed) to examine if they were different with candidates of distinct proficiency groups. Table 6 shows that off-list words (t= 2.218, p<.05) distinguished between candidates of two different proficiency groups. However, the differences were not statistically significant for the other lexical measures. The differences between the two groups of the variables types and tokens were not far off significance.

Table 5
Comparison Of Interviewer's And Candidates' Vocabulary Production

		Group A	Group B	Total
Types	Candidates	161.40	89.20	125.30
	Interviewer	105.20	85.70	95.45
Tokens	Candidates	377.10	171.10	274.10
	Interviewer	220.90	159.30	190.10
TTR	Candidates	.46	.53	.49
	Interviewer	.50	.54	.52
D	Candidates	67.88	60.05	63.96
	Interviewer	59.02	57.57	58.29
PLex	Candidates	.68	.34	.51
	Interviewer	.45	.30	.37
K1	Candidates	90.93	92.35	91.64
	Interviewer	93.88	94.61	94.24
K2	Candidates	3.91	3.82	3.87
	Interviewer	3.42	3.82	3.62
Off-lists	Candidates	5.15	3.83	4.49
	Interviewer	2.92	1.67	2.29

Table 6
Comparison of Interviewer's Lexical Variable Between Group A and B

Statistics							
	t-test for Equality of Means						
						95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Types	1.735	18	.100	19.50	11.24	-4.107	43.107
Tokens	1.887	18	.075	61.60	32.64	-6.975	130.175
TTR	-1.661	18	.114	-.04	.02	-.091	.011
D	.366	18	.718	1.45	3.96	-6.861	9.759
Plex	1.641	18	.118	.14	.09	-.040	.328
K1	-1.117	18	.279	-.74	.66	-2.120	.648
K2	-.903	18	.378	-.40	.44	-1.334	.532
Off-lists	2.218	18	.040	1.25	.56	.066	2.432

Figure 2 illustrates correlations of lexical measures between the interviewer and the candidates. P_Lex (r=.520, p<.05) and off-lists (r=.562, p<.01) were significantly correlated (two-tailed Pearson correlation). This indicates that the more rare words the candidates used, the more sophisticated the interviewer's vocabulary was.
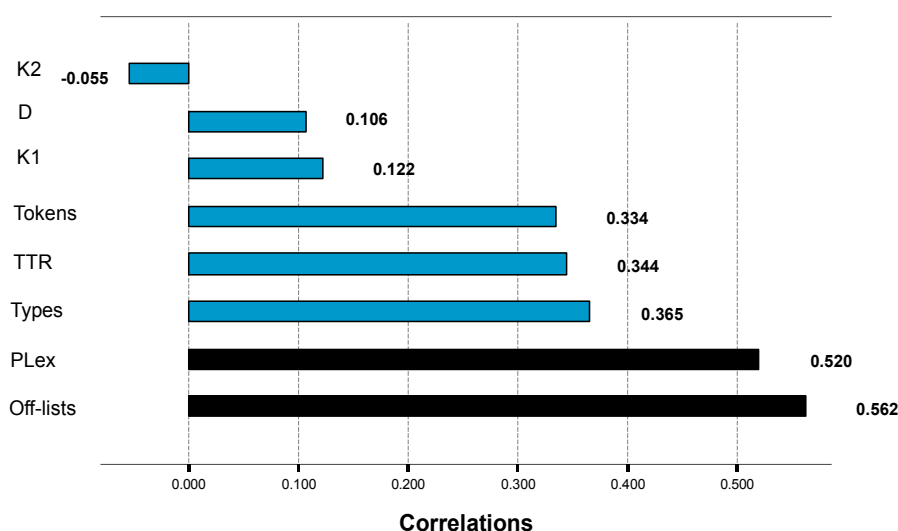

Figure 2. Correlations between paired lexical measures between the interviewer and the candidates (n=20)

Table 7 sets out Spearman correlations between candidates' vocabulary scores and IELTS scores and measures of the interviewer's vocabulary production. The results were statistically significant for P_Lex and off-lists with both vocabulary scores and IELTS scores. So, measures of interviewer's lexical sophistication seemed to be sensitive to both candidates' language proficiency levels and their lexical sophistication in the interviews (Figure 2).

Table 7
Correlations of the Interviewer's Vocabulary Measures and Candidates' Language Proficiency and Vocabulary Scores (n=20)

		Types	Tokens	TTR	D	Plex	K1	K2	Off-lists
Vocabulary	rho	.029	.046	-.103	-.033	.472*	-.432	-.019	.553*
	Sig. (2-tailed)	.903	.847	.664	.890	.036	.057	.938	.012
IELTS	rho	.207	.180	-.199	.135	.541*	-.348	-.134	.517*
	Sig. (2-tailed)	.382	.448	.400	.569	.014	.133	.573	.019

*. Correlation is significant at the 0.05 level (2-tailed).

Discussion

Candidates' Vocabulary Production and Scores

This section discusses vocabulary scores from two perspectives: (1) the relation between candidates' lexical statistics and vocabulary scores, and (2) the factors that may have affected the examiner's judgements on vocabulary in the OPIs apart from candidates' vocabulary production.

Most lexical measures of the candidates, such as types, tokens, P_Lex and off-lists (Table 3) statistically distinguished between candidates at different proficiency levels. However, D and TTR failed to have significant effects although they were not far from significance. Since TTR has been argued to be a function of sample size (Malvern and Richards 1997, 2002; Jarvis, 2002), the variation in text length in my data might be a cause for the insignificant difference. This finding underscored the contention that TTR was not a suitable measure for assessing texts of different lengths in oral contexts.

Regarding D, the insignificant correlation obtained in this experiment is not really surprising. Like TTR, D is also a type/token measure. Although most validation studies of the measure (Jarvis, 2002; Malvern and Richards, 1997, 2002; Malvern, Richards, Chipere, and Durán, 2004) so far have suggested that it is a valid measure of lexical diversity, D may be a function of text length. Owen and Leonard (2002) found children's (42 to 71 months) D values were significantly higher for 500-token speech than 250-token speech. Similarly, McCarthy and Jarvis (2007) cautioned that D did not work well with short texts (<100 tokens) or longer texts (>400 tokens), which implied that texts of various lengths might not be strictly comparable. Furthermore, the majority of the significant results of D were calculated using children's language samples (Klee, et al, 2004; Malvern, et al., 2004; Miralpeix, 2006). With respect to how useful D is at distinguishing between L2 candidates at different proficiency levels, the findings so far have not shown many significant results. Malvern and Richards's (2002) study suggested the D values correlated with language proficiency, but did not demonstrate that D values were significantly different between two proficiency groups in their study. Lorenzo-Dus and Meara (2005) reported that D did not distinguish between candidates at four proficiency levels. Likewise, Jarvis (2002) found that D values were not significantly different between students of grades 5, 7, and 9.

In addition to the insignificant effect by proficiency groups obtained in my experiment, D had only a moderate correlation with candidates' language proficiency (Table 5). The correlation (rho= .473, P<.05) was comparatively weaker than that of types, tokens, P_Lex and off-lists. Although current literature seems to suggest that D is a valid index of lexical diversity, measuring vocabulary with D should still be treated with caution in the context of oral examinations. Many questions about the measure remain unknown: how sensitive D is to measure vocabulary growth? How reliable is D in representing candidates' vocabulary knowledge at different proficiency levels? What is the impact of various oral tasks on this measure? Despite the validity argument over D, maybe it is time to reconsider how informative type/token lexical measures are in spoken vocabulary assessment.

While type/token based measures of lexical diversity appeared to be less promising in differentiating between proficiency levels in this experiment, list-based measures demonstrated a clear advantage over the other lexical measures. Both P_Lex and off-lists displayed significant effects between proficiency groups. In addition, both measures correlated significantly with candidates' overall language proficiency (IELTS scores, Table 4). Note also that P_Lex had the strongest correlation with IELTS among all lexical measures examined, implying that P_Lex might be effective in predicting overall proficiency levels in OPIs.

In addition, measures of lexical sophistication seem to correlate more strongly with vocabulary scores than with vocabulary output (tokens) and lexical diversity (types and D). P_Lex had the strongest correlation with the candidates' vocabulary scores (rho= .693, p<.01, Table 4). The results imply that the candidates' use of difficult words corresponded to higher vocabulary scores.

Adding to the statistical evidence, the examiner's definition for 'range of vocabulary' demonstrates explicitly the important role of lexical sophistication in her criterion for assessing vocabulary: 'the variety of vocabulary use, such as common use vocabulary to more specific vocabulary' (Table 1). Moreover, the interviewer's comments after the interviews further reveal that lexical sophistication has played a significant role

in her assessment of vocabulary. Frequent comments were made on candidates' using general expressions such as 'the vocabulary is basic' or 'the vocabulary is bad.' When asked to give specific examples to support such comments, the interviewer frequently mentioned the actual absence of rare words. Still, she did not provide any specific examples.

Note also that vocabulary scores were expected to correlate more strongly with the indices of lexical richness. Instead, they correlated more strongly with OPI scores and IELTS scores than vocabulary measures such as D and P_Lex (Table 4). Although the difference in the correlation values is not big, another question arises pertaining to the reliability of vocabulary scores in OPIs: is vocabulary judged as a discrete construct or from a holistic approach on the basis of general impressions on candidates' oral performance? This merits further research.

The Impact of Candidate-Interviewer Interactions on Candidates' Lexical Statistics and Vocabulary Scores

Besides statistical evidence on the interviewer's accommodation to candidates at different proficiency levels, the examination of the transcripts also revealed the effect of candidate-interviewer interactions on candidates' vocabulary output. For example, candidate Dai (medium ability) had the second highest P_Lex value in group B. A close examination of her transcripts (Example 1) shows that Dai's recounting of the words that the interviewer used might be a likely reason for her relatively high off-lists and P_Lex value. Such use of her words appear in bold in Example 1.

Example 1: (3-13)
  1 *DAI: yeah these **pictures** are very nice.
  …
→ 2 *INT: we showed all **photos** to every+body.
  3 *DAI: which one is you like?
  4 *INT: the one i took?
  5 *DAI: yeah.
  6 *INT: uh huh this one # and this one # this one.
  7 *DAI: yeah.
  8 *INT: it is not so good [=! laughs].
  9 *INT: yeah they are very old,, aren't they?
→ 10 *DAI: [>] <yeah they are very good **photos**>.
  11 *INT: yeah they are very good # my friend-: # visited some monks # who were

       having a big meeting out+side.
  12 *INT: it was amazing.
  13 *DAI: yeah.
→ 14 *INT: but he is a very good **photo+grapher**.
  15 *DAI: uh huh.
  16 *INT: it is just my hobby [=! laughs].
  17 *DAI: it is not bad # and this one is good.
  18 *INT: yeah that one is difficult.
  19 *INT: i did not take that one [=! laughs].
  20 *DAI: yeah.
  21 *INT: yeah this one is great.
→ 22 *DAI: your friend is a great **photo+grapher**.

When Dai was required to elicit information from the interviewer during Part 3 of the interviewer frame (Table 1), Dai asked about the photos hanging on the wall in the interviewer's office. She first used 'picture' (line 1), but changed to 'photos' (line 10) after the interviewer used that word (line 2). Dai's use of 'photographer' (line 22) is another instance of her recounting the interviewer's word (line 14) during the OPI. Since recounting on this occasion was not an instance of immediate repetition, it was included as Dai's productive vocabulary, which in turn, affected her off-lists and P_Lex value. Thus, the interviewer's vocabulary seemed to have an impact on candidates' vocabulary production, and hence accounted for 'superficially' high P_Lex values and more off-list words in their production.

Apart from recycling some of the interviewer's lexical production, candidates' vocabulary production in general was found to be an index of candidate-interviewer interactions, in particular of the interviewer's accommodation strategies. This section discusses whether the interviewer accommodated to candidates in terms of vocabulary use and if so, whether vocabulary scores were affected by such an accommodation.

The comparison between the interviewer's lexical measures and those of candidates revealed that the interviewer's overall lexical performance was at a lower level than that of the candidates, particularly with group A candidates (Table 6). Also, she used significantly more low frequency words (off-lists, Table 7) with candidates of high ability (group A) than with candidates of medium ability (group B). In addition, the interviewer's lexical measures showed significant correlations with candidates' P_Lex and off-lists, suggesting a relationship between the interviewer's vocabulary production and candidates' vocabulary use. This relationship may be discussed in terms of interviewer accommodation.

Evidence of the interviewer's accommodation can be observed from the questions she asked to the candidates in the OPIs. The interviewer's deviation from the interlocutor frame was prevalent in all 20 interview scripts. A preliminary examination on the interviewer's questions seems to suggest that she has adjusted the questions to the candidates' different proficiency levels.

Many researchers have found accommodation played a significant role in OPIs. Interviewers' adjustments in language use range from lexical simplification, grammatical simplification and comprehension check to a slowing down of speech rate, over-articulation, and ways of displaying questions (Ross and Berwick, 1992; Lazaraton, 1996). The correlations between the interviewer's and candidates' P_Lex and off-lists in the present study were significant. This indicates that the interviewer adjusted the 'difficulty' of her vocabulary both to the candidates' language proficiency levels and to the particular 'difficulty' level of their vocabulary in the OPIs.

However, the statistical analysis conducted also showed that the interviewer's P_Lex and off-lists correlated significantly with many measures of group A candidates, but not with those of group B candidates, with the single exception of interviewer's and candidates' off-lists. This means that the interviewer used her vocabulary differently with candidates of different proficiency groups. Similarly, Ross and Berwick (1992) found that interviewer's accommodation differed with the particular proficiency level of candidates. The results of my experiment, therefore led me to concur with their claim that accommodation 'reflected the interviewer's attempts to facilitate the communication of information during the process of the interview' (Ross and Berwick, 1992, p. 164).

In contrast to the results that the interviewer's P_Lex and off-lists were more sensitive to candidates' corresponding measures, Malvern and Richards (2002) found that lexical diversity responded most to candidates' language production. The interviewers in their study had notably lower D values than those of the candidates, especially for candidates of higher proficiency group. Such statistical differences of D prompted Malvern and Richards to suggest the presence of interviewer accommodation regarding D, even to suggest interviewers' over-accommodation to candidates of higher language proficiency. The disparity between the findings in the present study and Malvern and Richards' (2002) study confirms inter-examiner variation in OPIs.

One important consequence of the interviewer's accommodation is that it seems to have had an impact on candidates' vocabulary scores (Lorenzo-Dus and Meara, 2005; Van Moere, 2006). Significant correlations

were found between the interviewer's P_Lex and off-lists values with her judgements on vocabulary in my study. Similarly, Brown (2003) further argues that candidates are perceived to be more 'competent' when paired with an easy or teacher-like interviewer, one who is prone to accommodate his/her language to that of the candidates. In her view (2003, p. 1), 'the interviewer is implicated in the construction of candidate proficiency.' However, the extent of the impact of accommodation on scores, or in which aspects candidates' performances are influenced by accommodation, has to date not yet been explicitly spelt out, at least when vocabulary scores are concerned.

Conclusion

Drawing upon data from twenty OPIs in this experiment, I have explored candidates' vocabulary production, especially vis-à-vis their overall language proficiency, and the relationship between their vocabulary production and the vocabulary scores they were assigned. The analysis conducted in the experiment has yielded a number of significant results. For example, candidates' spoken vocabulary profiles in OPIs were found to differ by proficiency groups. In particular, high proficiency candidates produced more low frequency words and fewer high frequency words than those of candidates of medium ability. In addition, significant correlations were found between some of candidates' lexical measures and the interviewer's lexical measures and vocabulary scores she assigned, confirming that vocabulary is one important predictor of oral proficiency.

Regarding the various vocabulary measures, PLex shows potential to be an effective tool to differentiate oral vocabulary of the candidates in this experiment. In contrast, D, the popular measure of lexical diversity, failed to discriminate candidates' vocabulary output, which raised the question of D' s sensitivity as a measurement of lexical diversity for oral texts.

This experiment has, therefore, answered some of the questions I posed, while it has also opened up some areas to explore. Since OPIs are claimed to test communicative ability, it is reasonable to assume that assessment of interactive tasks (e.g. OPIs) and of monologic tasks (e.g. story-telling) may not be based on the same criteria. Thus, the results obtained in this experiment can only explain the approach of this particular examiner to her vocabulary judgement for OPI tasks. They cannot account for vocabulary assessment in other kinds of oral examinations; for example, the assessment of vocabulary in monologic tasks such as free discussion and story-telling tasks.

Finally, one important finding of this experiment is that the interviewer in this experiment displayed variation in her accommodation strategies with candidates of different proficiency, and her accommodation was found to have factored in her assessment of candidates' vocabulary despite the use of an interlocutor frame. For example, the interviewer's P_Lex and off-lists correlated significantly with candidates' P_Lex and off-lists, as well as with their vocabulary scores. These significant results further point out that the interviewer accommodation may pose a threat to the reliability of vocabulary scores in the OPIs. A preliminary analysis of both interviewer and candidates' discourses provide a straightforward example of the interviewer's influence on candidates' vocabulary production. Moreover, the example reveals the setback of using statistics alone in investigating accommodation issues: candidates' lexical statistics may be superficially high due to their recounting of the interviewer's lexical output. It is important, therefore, to explore the extent to which interviewers' accommodation has affected candidates' vocabulary from a complementary qualitative perspective. By doing so, a fuller understanding of vocabulary scores in the OPIs may be gained. This will be addressed by another article.

References

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests.* Oxford, UK: Oxford University Press.

Bell, H. (2003). *Using Frequency Lists to Assess L2 Texts.* Swansea University: Unpublished Thesis.

Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing, 20*(1), 1-25.

Brown, A., Iwashita, N., & McNamara, T. (2005). *An examination of rater orientations and test taker performance on*

English for academic purposes speaking tasks. Princeton, NJ: Educational Testing Service.

Cobb, T. (2007). The Web Vocabulary Profiler (WebVP), v1.5, from http://www.lextutor.ca/

Daller, H., Hout, V., R, & Treffers-daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24, 197-222.

Daller, H., & Xue, H. (2007). Lexical richness and the oral proficiency of Chinese EFL students. *Modelling and assessing vocabulary knowledge*, 93-115.

Duran, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics*, *25*(2), 220-242.

Engber, C. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*.

Espinosa, S. (2005). Can P_Lex accurately measure lexical richness in the written production of young learners of EFL? *PORTA LINGUARUM, 4*, 7-21.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: how distinct? *Applied Linguistics*, *29*(1), 24-49.

Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing, 19*(1), 57-84.

Jin, T., & Mak, B. (2013). Distinguishing features in scoring L2 Chinese speaking performance: How do they work? *Language Testing, 30*(1), 23-47.

Klee, T., Stokes, S. F., Wong, A. M. Y., Fletcher, P., & Gavin, W. J. (2004). Utterance length and lexical diversity in Cantonese-speaking children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 47*(6), 1396.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*(3), 307-322.

Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning, 48*(3), 365-391.

Lazaraton, A. (1996). Interlocutor support in oral proficiency interviews: The case of CASE. *Language Testing, 13*(2), 151-172.

Lim, G. S., & Galaczi, E. D. (2010). Lexis in the assessment of speaking and writing: An illustration from Cambridge ESOL's General English tests. *Research Notes, 41*, 14-19.

Lorenzo-Dus, N., & Meara, P. (2005). Examiner support strategies and test-taker vocabulary. *IRAL, 43*(3), 239-258.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd Edition ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Malvern, D. D., & Richards, B. J. (1997). Validation of a new measure of lexical diversity. In A. Ryan & A. Wray (Eds.), *Evolving models of language. Papers from the Annual Meeting of the British Association of Applied Linguists held at the University of Wales, Swansea, September 1996* (pp. 58-71). Clevedon, UK: Multilingual Matters.

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing, 19*(1), 85-104.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development.* New York, NY: Palgrave Macmillan.

McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing*, *24*(4), 459-488.

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp. 35-53). Cambridge, UK: Cambridge University Press.

Meara, P. (1999). The vocabulary knowledge framework. *Vocabulary Acquisition Research Group Virtual Library*.

Meara, P. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics, 26*(1), 32-47.

Meara, P., & Bell, H. (2001). P-Lex: A simple and effective way of describing the lexical characteristics of short L2 tests. *Prospect, 16*(3), 5-19.

Meara, P., & Milton, J. (2003). X_Lex: The Swansea vocabulary levels test: Newbury, MA: Express.

Meara, P., & Miralpeix, I. (2004). D_Tools. Swansea: Lognostics (Centre for Applied Language Studies, Swansea University).

Miralpeix, I. (2006). Age and vocabulary acquisition in English as a Foreign Language (EFL). *Age and the rate of foreign language learning. Clevedon: Multilingual Matters*, 89-106.

Miralpeix, I. (2007). Lexical Knowledge in Instructed Language Learning. *IJES, 7*(2), 61-83.

Morris, L., & Cobb, T. (2004). Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees. *System, 32*(1), 75-87.

Muncie, J. (2002). Process writing and vocabulary development: comparing Lexical Frequency Profiles across drafts. *System, 30*(2), 225-235.

Muñoz, C. (Ed.). (2006). *Age and the Rate of Foreign Language Learning.* Clevedon, UK; Buffalo, NY; Toronto, Canada: Multilingual Matters Ltd.

Nation, I. S. P. (2002). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.

Nation, P. (2007). Fundamental issues in modelling and assessing vocabulary knowledge. In H. Daller, J. Milton & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 288). Cambridge, UK: Cambridge University Press.

Nakatsuhara, F. (2008). Inter-interviewer variation in oral interview tests. *ELT Journal*, 62*(3)*, 266-275.

Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review/ La Revue canadienne des langues vivantes, 56*(2), 282-308.

Owen, A. J., & Leonard, L. B. (2002). Lexical Diversity in the Spontaneous Speech of Children With Specific Language Impairment: Application of D. *Journal of Speech, Language, and Hearing Research*, *45*(5), 927-937.

Read, J. (2000). *Assessing vocabulary*. Cambridge, UK: Cambridge University Press.

Read, J. (2005). Applying lexical statistics to the IELTS speaking test. *Research Notes, 20*, 12-16.

Richards, B., & Malvern, D. (2007). Validity and threats to the validity of vocabulary assessment. In H. Daller, J. Milton & J. Treffers-Daller (Eds.), *Modelling and assessing lexical knowledge*. Cambridge, UK: Cambridge University Press.

Richards, B., & Malvern, D. (2000). Accommodation in oral interviews between foreign language learners and teachers who are not native speakers. *Studia Linguistica, 54*(2), 260-271.

Richards, J. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 77-89.

Ross, S. (1992). Accommodative questions in oral proficiency interviews. *Language Testing, 9*(2), 173-185.

Ross, S., & Berwick, R. (1992). The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition*, *14*(2), 159-176.

Ross, S. (2007). A comparative task-in-interaction analysis of OPI backsliding. *Journal of Pragmatics, 39*(11), 2017-2044.

Schmitt, N., & Meara, P. (2000). Researching vocabulary through a word knowledge framework. *Studies in Second Language Acquisition, 19*(01), 17-36

Silverman, S., & Ratner, N. B. (2002). Measuring lexical diversity in children who stutter: application of vocd. *Journal of Fluency Disorders, 27*(4), 289-304.

Tidball, F. O., & Treffers-Daller, J. (2008). Analysing lexical richness in French learner language: what frequency lists and teacher judgements can tell us about basic and advanced words. *Journal of French Language Studies, 18*(Special Issue 03), 299-313.

van Hout, R., & Vermeer, A. (2007). Rich and happy? Reliability and validity of lexical richness measures. In H. Daller, J. Milton & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 93-116). Cambridge, UK: Cambridge University Press.

Van Moere, A. (2006). Validity evidence in a university group oral test. *Language Testing, 23*(4), 411-440.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing, 17*(1), 65-83.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics, 22*(02), 217-234.

Vermeer, A. (2004). The relation between lexical richness and vocabulary size in Dutch L1 and L2 children. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 173-189). Amsterdam, Netherlands; Philadelphia, PA: John Benjamins.

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review, 53*(1), 13-40.

Young, R. (1995). Conversational styles in language proficiency interviews. *Language Learning, 45*(1), 3-42.

Young, R., & Milanovic, M. (1992). Discourse variation in oral proficiency interviews. *Studies in Second Language Acquisition*, *14*, 403–424.

Yu, G. (2009). Lexical diversity in writing and speaking task performances. *Applied Linguistics*, 1-24.

The Value of Feedback and Conferencing in the Process Approach to Writing for Filipino and Thai Students in Higher Education: A Comparative Analysis

Analiza Liezl Perez-Amurao
(analizaliezl.amu@mahidol.ac.th)
Mahidol University International College, Thailand

Abstract

Using two groups of university students who enrolled in English 102, one from the Philippines and one from Thailand, this study looked into the role feedback and conferencing played in the process writing approach of the respondents. This study used a descriptive-qualitative causal design, adopting Dana Ferris' 1995 research in multiple-draft composition classrooms done at the California State University in Sacramento. To suit the specific needs of this study, Ferris' questionnaire was modified. Survey results were tabulated and questions placed under ordinal categories were averaged using SPSS. Results were analyzed using Categorical Data Analysis (CDA). Other questions were dealt with using a qualitative survey data analysis, a process of systematically searching for and arranging the answers to open-ended questions. This study underscores four major findings. First, students preferred both the quantitative and qualitative forms of assessment of their essays. Second, students preferred the process approach to writing, relying on the benefits the model promised them. Third, earlier drafts were found to have had positive influence on the succeeding drafts and the final copy. Lastly, this study demonstrates that revision is central in the improvement of every essay.

Key words: feedback, conferencing, process approach to writing

Introduction

The context for academic writing serves as a very crucial factor that determines largely the learner's success notwithstanding the axiom that one learns to write by writing and that many second language (L2) writers eventually turn out as good writers by generating an ample amount of outputs over time. Said context includes as well the manner through which a teacher helps the learners achieve learning objectives "through a variety of intervention strategies available to the classroom setting" (Kroll, 2003, p. 115). As Silva and Matsuda (2002) pointed out, writing is something that is always "embedded in a complex web of relationships between writers, readers, the text and reality" (p. 253). In recognition of the constantly changing nature of academic writing, it can be truly said that, in reality, academic writing operates inside a system of an academic community whose members find meaning in building partnerships with those whom they share values or pursue similar academic quests.

In Southeast Asian classrooms, however, this is not the norm. Where the writing pedagogy is dominated by the product-centered paradigm, better known as the traditional approach to writing, most students have yet to make themselves familiar with both the practices and the beliefs inherent in academic writing in universities where English is the medium of instruction (Ballard & Clanchy, 1991). Whereas students in the academic discourse communities are made to fulfill different writing objectives in different content areas, they are expected to carry out different roles as well. Given this, students nowadays are confounded with a problem in trying to meet the purposes and roles expected of them.

Batin (2003) maintained that what is difficult, however, is that with the dominance of the product-

centered paradigm, students are not given the chance "to discover, explore, and experiment on their linguistic repertoire and deprives them of a chance to plan, review, and reflect on their ideas" (p. 25). She further argued, "If writing were to develop and sharpen learners' mental acuity as well as communicative skills, it should allow for research, planning, drafting, revising, monitoring, feedback, and assessment" (2003, p. 32). Simply put, the process-oriented approach promises to be the better alternative to the product-centered paradigm.

As propelled by the benefits process writing offers, a number of researchers have raised relevant questions. Polio (2003) volunteered three kinds, the first of which is "Simply asking what the process is like" (p. 47). The second set of questions deals with interventions, an example for which is the giving of feedback and conducting conferencing. The third "limited number of studies has looked at change in writers over time" (p. 48). Of the three sets of questions Polio recommended, the one on the value of interventions via feedback and conferencing to students' composition is the most relevant to modern-day academic needs.

This research subscribes to Reid's (1983) findings based on her "own teaching that illustrates how teacher commentary taken out of context can be seen to misrepresent the dynamics of the classroom" (p. 121). Hence this study greatly considers the call for future researchers to remember that "[b]ecause teacher commentary, student reactions to commentary, and student revisions interact with each other, research needs to look at all three simultaneously" (Goldstein, 2001, p. 86).

This study has sought to look into the following: (1) what kind/s of assessment students prefer in their essay writing; (2) how students respond to the use of the process approach to writing; (3) how significant earlier drafts are in relation to the succeeding and final drafts, and; (4) what role revision plays in the revision of each draft/ paper.

In its aim to respond to the students' needs in the writing pedagogy from across the disciplines, the Department of English of a private Philippine university, in support of the principles of the process approach to academic writing, spearheaded a campaign in the College of Arts and Sciences (CAS) promoting the value of revision, teacher feedback, and conferencing. The Department of English initiated said drive upon learning through a preliminary survey that teachers from other disciplines/faculties were complaining over the kind of written outputs [using the English language] students handed in. Likewise, the Department of English found out a discrepancy in the academic writing performance of students in the English language and content area subjects, respectively. Initially, content area teachers were not as satisfied as the language teachers were in terms of the quality of writing students demonstrated. Apparently, students handed in better [final] papers in their language classes, as opposed to the quality of papers they gave in their content area subjects. This could be attributed to the fact that the language teachers, as opposed to the content area teachers, subscribed to the merit inherent in giving feedback to students' compositions and in conferencing with the students, all of which were significant in requiring them to write and revise academic papers through process writing in a multiple-draft setting.

Believing that much academic writing takes place in content courses, the Department of English highly encouraged the other departments to engage in a similar approach to writing, persuading them to become advocates of the axiom, 'Every teacher a language teacher.' This simply meant that because academic writing in said University was considered central in most content area subjects, the Department of English believed that the use of the process writing approach would not only be significant and beneficial to the language teachers and their composition students, but also to content area faculty and their students. Given the results that came out of the preliminary study this author conducted in 2006, this current study subscribes to the same basic tenet. For this current study, however, the author replicates said similar study in the higher education Thai context. Results of the Philippine study were used as a point of reference to checking whether Thai students in the university level have the same, similar, or different preferences concerning feedback-giving and conduct of conferencing. (For a discussion on the selection of the respondents, please see the Research Design Section.)

Process Writing

Given the exceptional spread of the English language written modality as an essential tool in both the academic and corporate milieu nowadays, it is imperative that relevant information about the people who use it and how it is used be consulted.

Process Writing in Academic Context

Every writing teacher who is sincere in helping a student succeed in the composition class knows that the social context, i.e., the classroom, plays a very significant role. So much so that "some variable of learning behavior which has correlational potential with instructional treatment" (Candlin, 2001, p. 122) should be identified to promote better teaching and learning. For one, identification of the criteria used in evaluating an academic paper is seen central in the process. As discussed earlier, the assessment and evaluation of the academic papers in this study used the following scheme: the revision of the first draft covered general content, the second was concerned with organization, and the third draft centered on style, grammar, and language.

While the process-oriented approach has so much to offer in terms of aiding learners attain writing proficiency, a significant component of academic achievement, most writing classes in the Philippine classroom setting have not yet taken advantage of its use and benefits. For one, in a quasi-experimental study of Batin (2003), she disclosed how the writing pedagogy in said country is still dominated by the product-centered or traditional method. Quoting Zamel (1999), she described it not a remote case as most countries experience this as well. Despite the seeming inability of the traditional writing approach to address the specific concerns of the students, it continues to be the leading method employed. To better understand and appreciate what the process-oriented process has to offer, it is best to assess it in opposition to the traditional approach.

According to Lynch (1997), the traditional approaches to writing are meant to make students compose correct texts that emphasize grammatical, vocabulary, spelling, and punctuation accuracy. Connor (1987) pointed out that such kind of writing, which considers style as the most important element putting greater focus on the linear process, follows a conventional paradigm that is mainly product-centered. The problem with this model, however, is that many teachers and researchers disapprove of it because it does not give a full picture of what successful writers actually do (Raimes, 1988). Such traditional model does not allow writers room for repair of writing errors.

On the contrary, the process-based approach draws attention to a writing method that allows development of both organization and meaning (Pincas, 2001). Process writing, inclusive of invention strategies, multiple drafts, and formative feedback, considers revision central to the practice.

Some pioneering studies looked into the value of process writing, giving light to the development of its principles and features. In Batin's (2003) review of key studies which paved the way for the development of process approach to writing, she identified five major findings, namely: (1) Writing is non-linear and recursive with overlapping and independent stages, such as pre-writing or planning, composing or writing and revising (Hayes & Flower, 1980); (2) The process approach gives students two crucial supports in writing, namely, time for the students to try out ideas and feedback on the content of what they write (Perl, 1980); (3) Process writing is an enabling approach for it nurtures the skills with which writers work out their solutions to the problems they set themselves; they shape their raw material into a coherent message towards an acceptable and appropriate form for expressing it (White & Arndt as cited in Batin, 2003); (4) Process approach writing enables students to explore and analyze than to imitate it; it allows them to go through a cycle of writing activities which require them to use efficient writing strategies to come up with a product that is not pre-conceived (Batin, 2003); (5) The process approach enables students to find the writing process as one for discovering new ideas and new language forms to express these ideas (Raimes as cited in Batin, 2003).

Another important concern that needs greater attention is the role of writing instructors as well as content area teachers to help students achieve academic competence. Although this present study attempts to highlight the use and benefits of the process approach in the English language class, it also aims to accentuate the

fact that the role of content area teachers can never be denied in making this endeavor successful. For one, Batin (2003) argued that the use of process approach helps teachers identify the kinds of students they have in their own classrooms thereby giving them clues how to deal with them academically. She cited five cognitive learning styles that students normally exhibit (Ausubel & Hill as cited in Batin, 2003). She maintained that knowledge of the said learning styles helps avoid, if not totally reduce, a mismatch between the students' learning styles and the teacher's teaching styles. When accommodating the value of the five cognitive learning styles in using the process approach to writing, both the teacher and the students reap the benefits especially during the giving of feedback and when doing revisions. Once the teacher recognizes the learning styles of the students, the teacher is placed in a better position, knowing what kind of feedback can be best given to every learner and how the feedback can be best communicated. When the teacher is able to match his/her manner of handling feedback to the students, it is likely that revision is better facilitated as feedback is assumed to be geared towards addressing specific points.

Feedback

Feedback is one of the two vital components of revision. The importance it plays in the writing process is further contextualized by way of discussing the need for it and the form it takes. This idea is best captured in Leki's "Coaching from the margins: Issues in written response" when she remarked, "How best to respond to student writing is part of the broader question of how to create a context in which people learn to write better or more easily" (1991, p. 57).

Interestingly, the context in which academic writing can be best taken is a social and cultural phenomenon. While it is true that the act of writing is most commonly thought of as the result of a writer's effort to pool ideas cognitively, it has to be viewed as well as a social and cultural act because it "takes place within a context that accomplishes a particular purpose…that is appropriately shaped for its intended audience" (Hamps-Lyons & Kroll, 1997, p. 8). This view emphasizes the social aspects of writing, specifically underscoring how one learns to write in the academic setting. This issue is crucial because it touches on the manner written outputs should be produced.

More specifically, Chaudron (1998) stated that the communication process in the writing classroom is partly made possible by the key role feedback plays. Feedback, juxtaposed with the mere notion of error correction, is an unavoidable component of classroom transactions. That is, learners will always derive information about their performance based on the teacher's response, or lack of it.

On the part of the teachers, they, too, see the value of their feedback as reflected in their students' writing, in how students behave towards it, and in the students' language acquisitions, generally speaking (Cohen, 1987). Lynch (1997) argued that a teacher's giving of feedback creates a "sound psychological sense," erasing whatever doubts and problems students have during the writing activity (p. 156). MacFarlane (as cited in Chaudron, 1975) touched on this as well, underscoring how feedback acts as a motivating force freeing the students from the apprehension of mistaking corrections as failures. Srole (1997) added that "…immediate feedback supplies clues to language problems" (p. 106).

However, it must be remembered that the kind of feedback given to students plays a crucial part. Hillocks (as cited in Fathman & Whalley, 1991) concluded that "…focused feedback can have an effect on certain aspects of writing" (p. 166). Whereas general or fragmented feedback leads to the non-improvement of the paper that is due for revision, focused feedback helps in showing students how to revise their texts successfully. White and Arndt (1996) advised teachers to "[r]espond as a genuine and interested reader, rather than as a judge and an evaluator" (p. 125). Fathman and Whalley (1990) revealed that "nearly all researchers agree that attention must be paid to both content and form" (p. 180) as "77% improved the content of their writing" (p. 185) when they were given feedback on said component.

On the contrary, Ferris (2003) reported that initial L1 research on giving feedback revealed a disheartening picture. She found out that students appeared to be disregarding teacher feedback. Hillocks (1986) and Knoblauch and Brannon (1981) reported that regardless of how the written feedback was provided, it did

not come out successful, supporting the teachers' aim to help students improve their writing. However, while some of these initial findings seemed discouraging, this study aims at negating earlier claims against feedback, believing that other factors might not have been considered leading to such a disparaging result.

This study, instead, subscribes to Raimes' (1985) conclusion: "With context, preparation, feedback, and opportunities for revision, students at any level of proficiency can be engaged in discovery of meaning" (p. 229). This present study maintains that it can be made possible through another equally significant and relevant step: conferencing.

Conferencing

When discouraging results came out regarding the students' reactions to feedback, the next major movement in the chronicle of response to student writing has geared towards strategies that are considered hugely apt (Raimes, 1985). This eventually paved the way to what is commonly known now as teacher-student conferences or conferencing.

Also known as communicative interaction or teacher-student talk, conferencing refers to the teacher's act of intervention during the writing process. Normally, a one-to-one conversation between teacher and student, it is a short conference that lasts from 10-15 minutes, enabling the teacher to discuss with the student problems in the latter's work (Peñaflorida, 1988).

According to Kroll (2003), one advantage of conferencing is that it lets the teacher discover probable misinterpretations a student might have about a prior written comment on writing issues. It also erases traditional concepts of classroom instructions as the transfer of information from a teacher who is knowledgeable to a student who is passively learning. A technique that need not be limited to the discussion of a specific draft, it yields results [from individual discussions] that are very satisfying (Chaudron, 1988).

Additionally, conferences are believed to be essential because they permit students to be in command of the communication, make clear their teachers' reactions, and negotiate meaning. "Furthermore, conferences enable teachers to assess how students react to their feedback" (Shin, 2003, pp. 3-4).

In conferencing, Peñaflorida (1988) further suggested, students are met individually by the teacher, rounding "out the process of discovering the unique backing records and needs of students, especially the first conference" all in the context of evaluation and response to student writing" (p. 85). Consistent with said views about conferencing, Graves (as cited in Peñaflorida, 1988) proposed a basic pattern of the writing conference, namely: (1) The student comments on the draft, (2) The teacher reads or reviews the draft, (3) The teacher responds to the student's comments, and (4) The student responds to the teacher's response.

Hinged on a popular classroom practice called scaffolding, conferencing, above all is a course design that puts learners as the reference point for decision-making, both in terms of content of lesson and form and style of teaching. Achieved through consultation and negotiation between teacher and learners, its main point is to make the students appreciate that their work is their property. That through it, students experience how it is to make "real-life decisions as well as decisions about schoolwork," realizing that the "real-world communicative needs of learners…take center stage in goal-setting" (Peñaflorida, 1988, p. 78).

A Conceptual Framework of Process-Oriented Writing in Southeast Asia

In keeping with the process-oriented approach to [academic] writing, the conceptual framework in the Figure 1 adopts the same framework that the preliminary study, which was conducted in the first participating university in the Philippines, subscribed to.
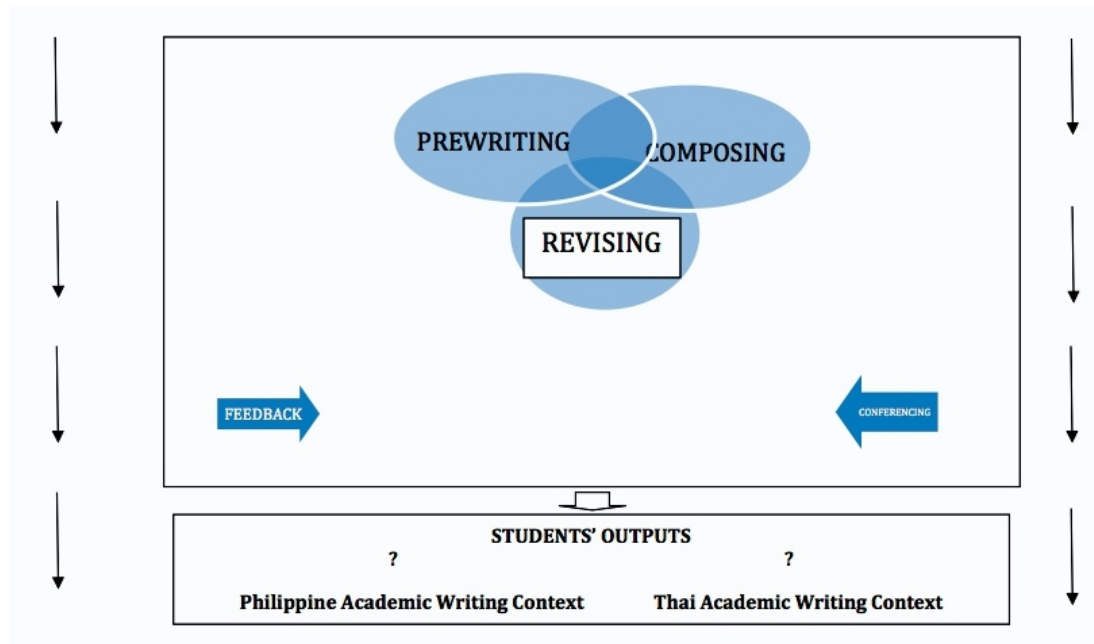
Figure 1. A Schematic Diagram of the Process-Oriented Approach with a Focus On Revision via Feedback and Conferencing

Consistent with the process every writer goes through, Figure 1 underscores the three most basic writing stages as represented by the three circles, namely the Pre-writing Stage, Composing/Drafting Stage, and Revising Stage, with emphasis on revision. While this study primarily focuses on the significant role revision plays in the composing process, both the discussion that follows and the schematic diagram above underscore the interdependence of the three stages of writing, as each phase does not work or operate in isolation. Revision being the focus of this study should be seen within the macro-framework of the process approach to writing.

As illustrated in Figure 1, this investigation aims to seek answers to questions primarily concerning the usefulness of revisions in the subjects' academic writing via the process approach. This study is anchored on the conceptual framework that argues that successful academic writing involves three major procedures, namely, prewriting, composing, and revising. Situated within the process approach context, prewriting, composing, and revising are believed to be both overlapping and independent stages (See Hayes & Flower, 1980); hence, the three separate yet overlapping circles representing the three steps. In the process of revising a paper, giving of feedback and holding conferencing sessions, which are all considered forms of teacher intervention, are key steps to producing satisfactory outputs. What this study ultimately seeks to find out is whether the key findings in the preliminary investigation that involved Filipino students from a Philippine private university would yield the same, similar, or different results when done with Thai students from a leading Thai university.

Research Questions

This study aimed to answer the following research questions:
1. Which kind/s of assessment do students prefer?
2. How do students respond to the use of the process approach to writing in their respective classes?
3. What is the significance of earlier drafts in relation to the students' production of the succeeding and final drafts?
4. What role does revision play in the revision of each paper/ draft?

The results of this study were conceived to benefit the following stakeholders in the future:

- Composition students, as findings of this study are to be used as baseline information to address classroom-specific needs in an effort to improve the current teaching strategies employed;
- Classroom teachers, both in the English language and content subjects, as (1) baseline information on the learners will help them understand the reason/s for, and remedy any discrepancy there is in the learners' academic writing performance, and (2) knowledge of the process writing approach and its use in the classroom will help improve the students' performance in the academic writing pedagogy;
- Body of language-area research, as relevant findings can be used to form part of the existing and latest explorations on said topic, aiming to enlighten researchers who are working and will work on the same research locale;
- Language curriculum planners, as results of this study can be made instrumental in devising language curricula, and;
- Instructional materials writers, as discussions and findings of this investigation can be helpful in designing teaching-learning aids in the academic writing pedagogy.

Methodology

This study made use of a descriptive-qualitative causal kind of research that focused on the importance of revision in process writing via feedback and conferencing in the [English] language class. As recommended by Ferris (1995) in her study, the coverage of this current investigation included administration of survey questionnaires, conduct of random interviews, analysis of composition outputs using Liz Hamp-Lyons' (1991) analytic and holistic marking systems, and keeping of journal entries.

Data Collection

This study made use of two sets of data. The first set of data came from a previous study done with participants who came from a private Philippine University. The second set of data came from participants from a leading Thai University. As this is a comparative analysis, the author purposely chose not to reveal the actual names of the students-participants and the Universities to protect their privacy. Actual results of said study, however, were used in all circumstances stipulated in and as required by the research design. Data collection was completed in 2012.

The first data site was a private Philippine University that emphasized the training of tertiary students via liberal arts education during the subject-students' first three years in school. Liberal Arts Education was offered by the University's College of Arts and Sciences, a service unit that offered general education courses. One of its courses, the English [language] course was founded on the unity of the trivium, namely, rhetoric, logic, and grammar. This course was offered by the Department of English, a department that handled [English] language courses. A separate department, the Department of Literature, offered literature courses [in English]. The second participating institution was a Thai state University that offered liberal education through its International College as its institutional strength. Just like the set-up in the first participating institution, this second participating institution offered English as a general education course taken up by its Freshman students. Said course was offered by the English Studies Program under the Humanities and Language Division. The EC2 course, from which the student-respondents in this study came, was the second credit course offered by the Program. Pre-requisites of this course included EC1, where students were made to write four essays using four different rhetorical patterns, and ERS, a non-credit course taken only by Freshman students whose language scores in the University Admission Exams did not satisfy the college's criteria. Should a student's entrance exam score satisfy the course placement requirement, they joined directly EC1 instead. This present study can be best appreciated within the context of the language courses, namely, English 102 for the Philippine university and Intermediate English Communication 2 (EC2) for the Thai university.

In the first participating institution's aim to promote and strengthen consciousness among its students the

spirit of unitas, the University credo, it [first participating institution] imbued its curriculum a singular focus on love for service and truth geared towards social development and attainment of national goals. This was made possible through the research topics the University's Department of English approved of. The topics for the English 102 papers were those that promoted social awareness among the students such as current issues and other socially relevant concerns that required the students' societal input.

The English Studies Program of the second participating institution, on the other hand, required topics that were theme-based. Set within the academic writing context, the writing of the students' papers was anchored on two broad themes, the first of which would have been pre-selected by the EC 2 teachers. The second theme was selected by the students. Similarly, EC2 primarily aimed at promoting awareness of social, political, cultural, economic, and religious issues among the students. In this investigation, as it was in the Philippine case, the focus was on the Thai students' writing of their academic argument essays.

However, teachers of both Philippine and Thai Universities recognized the need for the students to be abreast of issues that cultivated their social consciousness; hence the specific topics of the respondents from both Universities were student-nominated. The topics could be either "field-independent (more personal subjects)" or "field-dependent (academic and scholarly discourses)" (Batin, 2003, p. 28). Using the process approach within the academic writing context, the subjects were given the opportunity to explore topics of their choice to encourage them to be more motivated to conduct individual researches of their interest.

Additionally, consonant with the conditions that are typically required within an academic discourse community, both the Filipino and the Thai students were highly encouraged to produce research-driven papers. To advance this objective, the argumentative form of discourse was made a major course requirement. The argumentative form of discourse was a research-intensive rhetorical pattern. That is, apart from requiring students to follow strictly a convincing discourse structure, it called for a substantial discussion of the topic at hand, which normally could only happen when a student performed sound research. As indicated in the global scale used by Hamp-Lyons (1991) in her study of the academic writing proficiency of non-native students' written texts, the highest among the nine band descriptors that demonstrated good argumentative written discourse was the one that satisfied the readers fully. It demonstrated a very coherent organization, allowing the reader to follow the idea with ease. Appropriate points of view with complete and valuable supporting details were remarkably put forward, stating clearly the main ideas. There was also a clear and efficient correlation between the arguments and "the writer's experience or views" (1991, p. 147).

Participants

The first group was composed of 20 full-time Freshman students enrolled in the English 102 course of the university in the Philippines. Students in this group were homogenously categorized based on the University's sectioning system, the basis for which was the subjects' entrance exam results particularly the exam's English component. One of the main considerations in said selection was the homogeneity of the group. This was mainly because of the need to make a collective emphasis on the students' academic writing needs and other concerns that might not be possible if done with heterogeneously grouped learners.

The second group was composed of 19 full-time freshman students who were enrolled in the English Communication Skills 2 class of the Thai University. The homogenous grouping of the participants in this study was used as one of the criteria in the selection of the participants. The homogenous grouping was necessary to be able to direct a singular focus on the academic writing needs of the respondents not only in their language classes, but also in other content subjects. The homogenous grouping stemmed from the original sectioning system used by the participating university, based on the English-component results of the entrance examination of the subject respondents.

Both the Filipino and the Thai students were chosen as subjects in this comparative study based on the following reasons:

- Both groups were composed of Freshman college students and were enrolled in corresponding courses in

their respective universities.
- Both groups were taking their respective universities' second credit English language course that had the writing of the argument essay as the course's major requirement.
- Both groups were homogenously categorized in their respective universities based on each of their university's admission exam results, and both student groups came from the upper-middle socio-economic strata in the Philippines and in Thailand. It was assumed that based on the groups' socio-economic background, the students most likely had similar exposure in terms of, but not limited to, education, access to the English language within and outside the school setting, and others.
- Both the Filipino and the Thai students were expected to write about similar issues in their argument essays.

All participants of this study were full-time Freshman students of the two Universities. In the parallel study that was earlier initiated in the first participating university, a private university in Metro Manila, Philippines, the respondents were enrolled in an English 102. Their ages ranged from 16-18 years old, mostly coming from the upper-middle socio-economic strata. The class was of mixed genders.

In a replicate study done in Thailand, the respondents were of similar grouping. They were full-time Freshman students enrolled in English Communication Skills 2 class or EC2, the equivalent of English 102 in the first participating university. The students' ages ranged from 17-18 years old, and they were of mixed genders as well. Just like the Philippine-based respondents, the Thai students also came from the upper-middle socio-economic strata.

Purposive sampling was used in the Thai-university study, as it was the same sampling method used in the parallel study in the first participating university. Said sampling method had been chosen based on the existing students' admission protocol observed by both Universities. Both sets of respondents were homogenously grouped based on the sectioning system determined by their scores in the language component they obtained in the university entrance examination. The university entrance exam, primarily a placement type of test, had been instrumental in the selection of the participants in both the Philippine and Thai investigations, ensuring that the selected respondents performed homogenously based on their entry skills in the English language component. The types of learners subjected to this study belonged to the group whose scores in the University entrance exam did not qualify for English 101, in the case of the Philippine-based participating university, and English Communication Skills 1 or EC1, in the case of the Thailand-based participating university. In both cases, English 101 in the Philippines and English Communication Skills 1 or EC1 in Thailand were the first regular credit courses in the respective schools' English curriculum. Not having obtained the required score to qualify for the regular English course offerings of the Universities, the respondents needed to take the compulsory zero-credit English language courses, English Lab and ERS for the Philippine-and Thailand-based schools, respectively.

Instruments and Analysis

Learner Questionnaire

The questionnaire used in the original Ferris' (1995) study was modified to suit the specific needs and context of the current investigation. The modification to the questionnaire involved questions that called attention to the role revision via feedback and conferencing plays as an integral part of the typical academic writing process. The composition survey, a combination of qualitative and quantitative questions, covered approximately 23 writing issues. Eight questions called for answers under the ordinal category, e.g., "a lot", "a little", "some", etc. Said responses were statistically analyzed using the Categorical Data Analysis (CDA). The rest of the items drawing out subjective responses were treated using a qualitative analysis. Responses to the questions were clustered to establish and further examine the patterns based on the collected data.

Analysis of Composition Outputs

This study used a descriptive-qualitative causal design, underscoring the importance of revision in process

writing within an academic context. Considered one of the key skills every writer should develop, revision should be aided by feedback and conferencing. This study employed Dana Ferris' study done in 1995 at the California State University in Sacramento and included administration of Ferris' survey questionnaire that was slightly modified. It also included a modification Tajonera (2001) incorporated when she did her own descriptive study of students' attitudes about feedback. Similarly, the author of this study also made modifications to said instrument to suit the current study's objectives.

Consistent with one of the objectives of this investigation, the author looked into the progress the subjects of the current study demonstrated. However, unlike what was done in the preliminary investigation involving the first participating university, analysis of the composition outputs of the second participating university's respondents was based primarily on the quality of their outputs hinged on content, organization, and style, without using Hamp-Lyons' (1991) analytic (Appendix A) and holistic (Appendix B) marking systems. Said analytic and holistic marking systems were not used as the second participating university used a different system. Given such a modification, however, the author assumed that the most basic assumption this investigation was hinged on would be fulfilled.

Findings

In keeping with the research objectives cited in this study and to be able to arrive at a comprehensive comparative analysis between the two identified subject-groups, namely, the Filipino and the Thai college students, Questions 9, 13, 14, 18, 19, 20, and 22, which came with the original survey questionnaire, were excluded from the findings, discussion, and conclusion. This, in effect, resulted in the use of only 16 most relevant survey questions discussed in the succeeding pages.

The participants' responses to Question 1A ("Are you aware of the process writing approach?") indicates that all of the 20 Filipino respondents already had an idea about the process approach to writing before taking up English 102. The process approach was introduced at different stages in the students' tertiary-level academic life for the most part. An exception to this was one response from a Filipino student who claimed to have learned about this approach from as early as high school. Similarly, all of the 19 Thai survey respondents confirmed that they already knew about the process writing approach even before they reached EC 2.

The participants' response to Question 1B ("Did you find it helpful when you first used it in writing?") demonstrates that all of the Filipino and the Thai students found the process writing approach useful when they first used it in their respective writing classes.

The participants' responses to Question 1C, ("Do you find it helpful now in your English 102/ EC2?") show that a little more than half of the Filipino respondents (60%) stated three major reasons why they all found the process helpful. On the other hand, this question reveals that majority of the Thai respondents (74%) found it useful when developing the paper and organizing their ideas. Some said it allowed them to see their mistakes. Others claimed that the process approach made them more aware of the required structure of the paper they had to produce. The respondents' general commentary centered on the valuable assistance the approach gave them, allowing them to do a better revision.

Table 1
Responses to Question 2: "How many of each composition do you read over again when your instructor returns it to you?"

	All of it	Most of it	Some of it	None of it
		Filipino students		
1st/3rd drafts	60%	30%	10%	-
Final drafts	70%	20%	10%	-
		Thai students		
1st/3rd drafts	52.63%	42.10%	5.26%	-
Final drafts	63.15%	31.57%	5.26%	-

As shown in Table 1, the participants' responses to Question 2 ("How many of each composition do you read over again when your instructor returns it to you?") indicate that a majority of both Filipino and Thai students claimed to have read returned preliminary drafts (1st to 3rd), although a small percentage (30%) of Filipino respondents read only some of the final drafts, whereas 10% did not read the returned final draft at all.

Table 2
Responses to Question 3: "How many of your instructor's comments and corrections do you think about carefully?"

	All of it	Most of it	Some of it	None of it
	Filipino students			
1st/3rd drafts	60%	40%	-	-
Final drafts	60%	-	30%	10%
	Thai students			
1st/3rd drafts	57.89%	42.10%	-	-
Final drafts	63%	37%	-	-

For Question 3 ("How many of your instructor's comments and corrections do you think about carefully?"), as shown in Table 2, all of the Filipino respondents thought about the instructor's comments and corrections carefully. Similarly, all of the Thai respondents said the same concerning all the drafts. These results show that majority of the students from each group reflected upon their teachers' comments, proving that despite the slightly varying percentages, the process approach engaged most of the respondents in both groups even after the final draft had been written and marked.

Question 4 was about the content of the teachers' correction (Table 3). Of the five language writing criteria, namely, organization, content/ ideas, grammar, vocabulary, and mechanics, a consistent high number of comments and corrections for all the five language writing criteria were noted among the Filipino respondents. The Filipino students' responses spread out relatively thinly, concentrating highly on "all" and "some." On the other hand, the Thai respondents' comments and corrections were highest for content/ ideas and the lowest was on vocabulary and mechanics. Comments and corrections on organization and grammar were moderate. Answers to this whole range of questions resulted in a very wide spread among the Thai respondents.

For the final draft, the Filipino respondents claimed to have received similar amount of comments and corrections for all the writing criteria. The Thai respondents, on the other hand, claimed to have received comments and corrections on all the language writing criteria for their final drafts, much lower compared to those they got for their preliminary drafts. The only exception was for the comments and corrections on mechanics, which only dropped to .53%. For the Thai respondents, the degree of comments and corrections involving the final draft changed significantly in some writing needs, with more respondents claiming fewer comments and corrections. For the Filipino respondents, results similar to their responses involving the first to the third drafts came out. This may mean that the Thai respondents must have significantly reduced the errors from the preliminary drafts to the final one. On the other hand, it may be worth a further investigation why results concerning the preliminary drafts and the final draft did not differ that much or not at all among the Filipino respondents despite their claim that they read the returned drafts to them more than the Thai students did.

Table 3
Responses to Question 4: Students' Perception of Correction Content

	A lot	Some	A little	None	No Answer
	Filipino students				
1ˢᵗ /3ʳᵈ drafts					
Organization	70%	30%	10%	-	-
Content/ Ideas	70%	30%	10%	-	-
Grammar	70%	10%	10%	-	10%
Vocabulary	70%	10%	10%	-	10%
Mechanics	70%	10%	10%	-	10%
Final drafts					
Organization	70%	30%	-	-	-
Content/ Ideas	70%	20%	-	-	10%
Grammar	60%	10%	10%	-	20%
Vocabulary	60%	10%	10%	-	20%
Mechanics	60%	10%	10%	-	20%
	Thai students				
1ˢᵗ /3ʳᵈ drafts					
Organization	26.32%	31.58%	5.26%	21.05%	15.78%
Content/ Ideas	26.31%	42.10%	21.05%	5.26	5.26%
Grammar	21.05%	36.84%	21.05%	10.52%	10.52%
Vocabulary	5.26%	26.31%	36.84%	21.05	10.52%
Mechanics	5.26%	26.31%	31.57%	21.05	15.78%
Final drafts					
Organization	10.52%	21.05%	47.36%	10.52%	10.52%
Content/ Ideas	10.52%	26.31%	31.57%	21.05%	10.52%
Grammar	10.52%	26.31%	36.84%	21.05%	5.26%
Vocabulary	5.26%	10.52%	42.10%	36.84%	5.26%
Mechanics	5.26%	15.78%	21.05%	47.36%	10.52%

When asked how much attention was given to the first to third-draft comments and corrections (Table 4), the Filipino respondents said that the attention they gave to their preliminary drafts was consistently high for all the five criteria, except for one respondent who chose not to answer the question on grammar, vocabulary, and mechanics. In contrast, the Thai respondents claimed to have paid more attention to organization and content/ ideas the highest, followed by the one on grammar, leaving the least significant amount of attention to vocabulary and mechanics. The Thai respondents' degree of attention turned out to be high when correlated to the comments and corrections given by their instructor, especially on organization and content/ ideas, surpassing the responses of the Filipino students. They seemed, however, to believe that they did not think about said language writing areas as much as they did for grammar, vocabulary, and mechanics. The Filipino respondents, in contrast, generally had a consistent impression of the degree of attention they thought they gave to the different comments and corrections given them across the language writing needs.

Table 4
Responses to Question 5: Students' Reported Amount of Attention to Instructor's Comments and Corrections

	A lot	Some	A little	None	No Answer
			Filipino students		
1st /3rd drafts					
Organization	70%	30%	-	-	-
Content/ Ideas	70%	30%	-	-	-
Grammar	70%	10%	10%	10%	-
Vocabulary	70%	10%	10%	10%	-
Mechanics	70%	10%	10%	10%	-
Final drafts					
Organization	70%	30%	-	-	-
Content/ Ideas	70%	20%	-	-	10%
Grammar	60%	10%	10%	-	20%
Vocabulary	60%	10%	10%	-	20%
Mechanics	60%	10%	10%	-	20%
			Thai students		
1st /3rd drafts					
Organization	78.94%	21.05%	-	-	-
Content/ Ideas	78.94%	15.78%	-	5.26%	-
Grammar	52.63%	36.84%	10.52%	-	-
Vocabulary	36.84%	36.84%	15.78%	10.52%	-
Mechanics	36.84%	31.57%	15.78%	15.78%	-
Final drafts					
Organization	63.15%	21.05%	15.78%	-	-
Content/ Ideas	63.15%	26.31%	5.26%	5.26%	-
Grammar	31.57%	42.10%	10.52%	10.52%	5.26%
Vocabulary	42.10%	21.05%	15.78%	15.78%	5.26%
Mechanics	47.36%	21.05%	21.05%	10.52%	-

For the final draft, both the Filipino and the Thai groups claimed to have paid significant attention to organization and content/ ideas and a moderately significant attention to vocabulary and mechanics. Among the Filipino students, the attention they gave to organization and content/ ideas in the final draft was consistent with the one they gave when working on the first to the third drafts. The percentage, however, dropped slightly for grammar, vocabulary, and mechanics. The Thai respondents, on the other hand, gave the least amount of attention to grammar. For the Thai respondents, the attention they gave to the preliminary drafts dropped slightly for organization, content/ ideas, and grammar when compared to the final draft, but increased slightly for vocabulary and mechanics. These results reveal that writing students using the process approach would tend to lessen the amount of work exerted when dealing with the final draft.

Question 6 required the respondents to describe what they did after they read the instructor's comments and corrections ("Describe what you do after you read your instructor's comments and corrections (e.g., Do you look up the corrections in a grammar book? See a tutor? Rewrite your paper?"). A majority (60%) of the Filipino student-respondents said they re-wrote their papers. Others claimed to have done the following: re-evaluated their paper, looked up the corrections to make better revisions, and consulted peers and family members. Similarly, a majority (84.21%) of the Thai respondents said they did similar tasks such as the following: read again the essay, tried to understand the comments and corrections, and revised their own paper after. A little less

than half (42.1%) said that they looked for reasons/ information why they were wrong and then similarly revised their paper. A small percentage (21.05%) said that they sought the help of a friend or a higher-level student, revised their paper into the right format, and checked the APA style and other online materials. Other students' feedback also included the following: looked up words in the dictionary, used Google for correct vocabulary and more information, among others.

Question 7 ("Are there ever any comments or corrections that you do not understand? If so, can you give any examples?") inquired whether the respondents had a difficulty comprehending the instructor's comments and corrections. Among the Filipino respondents, a majority (70%) said "no." For the lone respondent who said "yes," the difficulty was linked to not being able to know exactly how "to expound on the paragraph" and being "a little bit confused with the marking process." Among the Thai respondents, a small number (26.31%) said "yes," whereas a majority (73.68%) said "no." The difficulty, they claimed, was due to the "simple" warnings stating that something was wrong, but no specific correction was given.

Question 8 looked into what the respondents did with corrections that they did not understand ("What do you do about those comments or corrections that you do not understand?"). Answers from both the Filipino and the Thai students revealed that both groups resorted to further enhancing their respective essays, at times, doing similar or same actions; that is, they sought help whenever and wherever help was available.

Question 10 asked if the instructor's comments and corrections helped the respondents improve their composition writing skills ("Do you think that your instructor's comments and corrections helped you to improve your composition writing skills? Why or why not?"). For the Filipino respondents, a majority (90%) said "yes," while one said "somehow." Those who said "yes" gave the following reasons: "It gave me the challenge to do better," "She [teacher] made me realize the errors that I have committed such as the limitations that should be observed in the content," "Because it made my writings better," "In a way, it helped me improve in writing compositions," and "Tried to avoid writing the same thing in my composition." The one who said "no" explained: "Because I only needed minor corrections in organization." All 19 of the Thai respondents said "yes." The most commonly cited reason from the Thai respondents was that the comments and corrections were "able to help them improve their writing, telling them where they went wrong." Others said that their instructor's ideas to help further improve the paper were "a lot better than mine." Other reasons were as follows: "It makes me think more carefully," "Comments show whether my paper is on the right track or not," "It helps me make my work more organized," and "These comments guide me how to move on with the paper." As a whole, both groups admitted to having improved their papers through the help of the process approach. Reasons were varied, but everything pointed to the immense paper improvement made via the process approach channeled through the instructor's giving of feedback and the conduct of conferencing.

In the last two major points in this discussion, findings for Questions 12, 15, and 16 have been grouped together, while findings for Questions 11, 17, 21, and 23 were combined as well to provide a more comprehensive analysis for items that are on similar research strands, hence the grouping of questions that thereafter follows.

The students' responses to Question 12 ("Which form of assessment do your prefer?") reveal that a majority of the Filipino and the Thai respondents (100% and 84.21%, respectively) preferred a form of assessment "with detailed comments and a numerical mark," a choice that was generally consistent with how they responded to the feedback and comments given by their instructor on their individual papers. Only some of the Thai students (10.52%) indicated a preference for only an overall numerical mark for the composition.

When asked to rate themselves as learners (Question 15: "How would you rate yourself as a learner?"), a majority of the Filipino respondents (90%) claimed themselves to be "good" as opposed to many of the Thai respondents (52.63%) who rated themselves as "fair." None of the students in the two groups rated themselves as "excellent" or "poor."

The respondents' self-rating concerning their performance as learners and as composition writers (Question 16: "How would you rate your skills in writing compositions?") was similar: A majority of the Filipino students (70%) and the Thai students (73.68%) considered themselves "good" writers, while none considered

themselves "excellent" or "poor." Interestingly, their self-rating correlated to the kind of revisions and attention they paid to the comments given by their instructor. The more attention the respondents gave to the feedback and comments given by the instructor, the higher was their assessment of themselves as learners and vice-versa.

Question 11 ("Do you think that your instructor's comments and corrections helped you discover and/or improve your skills in revising your paper?") checked on whether the instructor's comments helped them discover and/or improve their skills in revising their paper. Among the Filipino respondents, a majority (90%) said "yes," whereas one respondent did not answer the question. Those who said "yes" gave the following reasons: "It helped me change my writing style so it could be understood easier," "Because I knew what to replace or I knew where my errors were," "I was able to discover new styles in writing and improve my writing skills to make my composition easier to be understood by almost any type of reader," "It helped my paper to be more organized," "I got to know how to do revising already, on what was needed to be revised," "It did because I was able to increase my awareness of the writing process," and "Every detail was clarified." The Thai respondents, on the other hand, all said "yes." They said that improvements in their papers were made possible given the following reasons: "Without the instructor's comments, I would not know what is right or wrong with my paper," "I learned how to locate properly the thesis/ topic sentence in the essay," "It helped me know if the thesis statement and the topic sentence are related to one another," "I revised my paper accordingly so I would not repeat the same mistakes next time," "Makes it easy [for me] to find my own mistakes," "We need others to know whether they understand [us] or not," and "Reduced errors."

To Question 17 ("Generally speaking, I find the teacher's comments at the end of my paper helpful/ not helpful"), all the Filipino respondents (100%) and a majority of the Thai respondents (95.25%) indicated that they found their instructor's comments "very helpful."

Question 21 asked the respondents whether conferencing helped them clarify their thoughts/ ideas with their teacher ("Do you think conferencing helps you clarify your thoughts/ ideas with your teacher?"). For the Filipino respondents, a majority (90%) said "yes." They, however, did not give further reasons for their answer. On the other hand, all 19 of the Thai respondents confirmed that conferencing helped. Majority of the explanations centered on the value of conferencing in helping them improve their thesis statements, refine their ideas, and find a better focus.

Lastly, Question 23 asked the respondents if conferencing helped them discover and/ or improve their skills in revising their paper ("Do you think conferencing helps you discover and/ or improve your skills in revising your paper?"). Whereas all (100%) of the Filipino respondents said "yes," 94.73% of the Thai respondents said it did.

Discussion and Conclusion

What is worth noting is that the Filipino and the Thai respondents' replies to the last four questions correlated to their responses to earlier questions. The findings yielded in the last four remaining questions demonstrate one overarching result: That both groups found the use of the process approach to writing useful, ultimately helping them to make good revisions. This over-all response of both groups of respondents confirms the basic tenet which Zamel (1999), Ferris (2003), Reid (1993), among others, espouse. Both groups of students found the approach relevant and useful regardless of the differences in academic contexts within which the 2006 Philippine-based study and the 2012 Thailand-based investigation were conducted. Having said this, it only means that majority, if not all, of the respondents from both groups found and valued revision as a process that was central to good academic writing, irrespective of some specific pedagogical contexts within which the two groups of respondents were situated in. Indeed, in keeping with what Matsuda (2003) pointed out, both groups were in agreement that the use of feedback and conferencing in the process approach offered them a venue through which a better paper could be produced.

Results concerning corrections that involved the various language writing needs of the respondents indicated that while they were initially homogenously grouped based on their university admission results, the

Thai students' language writing needs covered almost all of the language writing criteria in varying smaller percentages, thus implying their highly heterogenous writing skills. The comments on and corrections to the Filipino respondents' papers, on the contrary, indicated a somehow homogenous set of writing skills. In this regard, it is interesting to find out in future researches whether this performance was influenced by their prior exposure to the process approach or not and/ or whether other factors interplayed that resulted in this specific finding.

It might be worthwhile to check further as well if the Filipino respondents' revisions went beyond what the comments and corrections required, making them come up with more new inputs that, in return, required again a new set of comments and corrections in the final draft. If this conjecture would hold true, the Thai students then can be comparatively said to be more conservative in their revisions, whereas the Filipino students could have been more aggressive. And if this was the case, this particular finding then corroborates with what Batin (2003) mentioned in reference to Ausubel and Hill's (as cited in Batin, 2003) cognitive learning styles. That is, learners can be classified differently as demonstrated through the different ways they perform in class. Nevertheless, this finding interestingly presents another topic for future researches.

Looking at the whole spectrum of answers from the two groups, one general commentary that can be made is that despite the varying ways the students responded to their instructor's comments and corrections, they all responded with a common end in mind. All of the respondents wanted to make a better revision of the preliminary drafts resulting in a final copy that was a product of various scaffolding activities gathered from the different people they had connections with who were either willing or available to help.

While majority of the respondents from both groups claimed that they understood the comments and corrections, those who had a hard time remarked that the instructor's comments and corrections were general ones, hinting that the students wanted specific, if not very specific, feedback. In this case, this study provides academic writing teachers a general idea about what can be possibly expected from their students, perhaps even giving them an advanced notice as to how they can address such an issue aligned with their own school's academic writing policies.

Although the surveys with the Philippine and the Thai students were carried out separately under some varying conditions, this study constantly upholds the significance of the assistance (or lack of assistance) of writing teachers to their students. In addition, findings of this study espouse the idea that there is always the attendant pedagogical merit in revising papers channeled through feedback and conferencing, if only to advance every student's interest.

In short, this study's findings suggest the following: First, students prefer both the quantitative and qualitative way of assessing their papers, that is, through detailed comments and a numerical grade. Second, students show preference for the process approach to writing anchored on the benefits the use of the model promises them, that is, improved writing skills and better grades. Third, in keeping with previous findings, earlier drafts always have positive influence on the succeeding drafts and the final copy. And fourth, revision has been proven to be central in the improvement of every paper.

References

Ausubel & Hill. (n.d.) In Batin, E. The process approach to writing: A principled and research-based writing pedagogy. *The ACELT Journal*, *7*, 25-32.

Ballard, B. & Clanchy, J. (1991). Assessment by misconception: Cultural influences and intellectual traditions. *Assessing Second Language Writing in Academic Contexts*. Norwood, NJ: Ablex.

Batin, E. (October 2003). The process approach to writing: A principled and research-based writing pedagogy. *The ACELT Journal*, *7*, 25-32.

Candlin, C. & Mercer, N. (Eds.). (2001). *English Language Teaching in its Social Context*: *A Reader*. USA: Routledge.

Chaudron, C. (1988). *Second Language Classrooms: Research on teaching and learning.* New York, NY: Cambridge University Press.

Cohen, A. (1987). Student Processing of Feedback on their Compositions: Teacher and Student Verbal Report. *Understanding Learning Styles in Second Language Classroom.* J. Reid (Ed.). Englewood Cliffs, NJ: Prentice Hall.

Connor, U. (1987). Research Frontiers in Writing Analysis. *TESOL Quarterly, 21,* 677-696.

Fathman, A. & Whalley, E. (1990). Teacher Response to Student Writing: Focus on Form Versus Content. *Second Language Writing: Research Insights for the Classroom.* B. Kroll (Ed.). New York, NY: Cambridge University Press.

Ferris, D. (1995). Student reactions to teacher response in multiple-draft composition classrooms. In *TESOL Quarterly, 29,* 33-53.

Ferris, D. (2003). Responding to writing. In *Exploring the Dynamics of Second Language Writing.* New York, NY: Cambridge University Press.

Goldstein, L. (2001). The social context for language learning: A neglected situation? In *English Language Teaching in its Social Context: A Reader*. New York, NY: Routledge.

Hamp-Lyons, L. (1991). *Assessing Language Writing in Academic Contexts*. NJ: Ablex Publishing Corporation.

Hamp-Lyons, L. & Kroll, B. (1997). Writing: Composition, community, and assessment. *TOEFL Monograph Series Report No. 5.* Princeton, NJ: Educational Testing Service.

Hayes, J. & Flower, L. (1980). *Academic writing in a second language: Essays on research and pedagogy.* D. Belcher & G. Braine (Eds.). New York, NY: Ablex Publishing Corporation.

Hillocks, G. Jr. In Fathman, A. & Whalley, E. (1990). Teacher Response to Student Writing: Focus on Form Versus Content. *Second Language Writing: Research Insights for the Classroom.* B. Kroll (Ed.). New York, NY: Cambridge University Press.

Hillocks, G. Jr. (1986). Responding to writing. In *Exploring the dynamics of second language writing.* B. Kroll (Ed.). New York, NY: Cambridge University Press.

Knoblauch & Brannon. (1981). Teacher commentary on student writing: The state of the art. *Freshman English News, 10,* 1-4.

Kroll, B. (Ed.) (2003). *Exploring the dynamics of second language writing.* New York, NY: Cambridge University Press.

Leki, I. (1991). The Preferences of ESL Students for Error Correction in College-writing Classes, *Foreign Language Annals 24 (3), 203-218.*

Lynch, T. (1997). *Communication in the language classroom.* Hong Kong, China: Oxford University Press.

MacFarlane, A. (1975). In Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning.* New York, NY: Cambridge University Press.

Matsuda, P. K. (2003). Second Language Writing in the 20th Century. *Exploring the dynamics of second language writing.* New York, NY: Cambridge University Press.

Penaflorida, A. H. (1988). Non-traditional forms of assessment and response to students' writing: A step towards learner autonomy. *Learners and Language Learning.* W. A. Renandya & G. M. Jacobs (Eds.). SEAMEO Regional Language Center. *Anthology Series, 39*.

Perl, S. (1980). Understanding composing. *College Composition and Communication*, *31(4),* 363-369.

Pincas, A. (2001). Structural linguistics and systematic composition teaching to students of English as a foreign language. *Landmark Essays on ESL Writing.* T. Silva & P. K. Matsuda (Eds.). Marwah, NJ: Hemagoras Press.

Polio, C. (2003). Second language writing research: What we investigate and how. *Exploring the Dynamics of Second Language Writing*. In B. Kroll (Ed.). Cambridge, UK: Cambridge University Press.

Raimes, A. (1985). What unskilled ESL students do as they write: A Classroom Study of Composing. *TESOL Quarterly 19(2)*.

Raimes, A. (n.d.). In Batin, E. The process approach to writing: A principled and research-based writing pedagogy. *The ACELT Journal*, *7*, 25-32.

Shin, S. J. (2003). The reflective second language writing teacher. *ELT Journal, 57(1)*, 3-10.

Srole, C. (1997). Pedagogical responses from content faculty: Teaching content and language in history. *The content-based classroom: Perspectives on integrating language and content.* M. Snow & D. M. Brinton (Eds.).

Whiteplains, NY: Longman.

Reid, J. (1993). In *Exploring the Dynamics of Second Language Writing*, B. Kroll (Ed.). New York, NY: Cambridge University Press.

Silva, T. & Matsuda, P. M. (Eds.). (2002). *Landmark Essays on ESL Writing*. NJ: Hemagoras Press.

Tajonera, L. K. (2001). What feedback I see and how I respond to feedback: A study of learner perceptions and responses to feedback. MA Thesis. Philippines: Ateneo de Manila University.

White, R. & Arndt, V. (n.d.). In Batin, E. The process approach to writing: A principled and research-based writing pedagogy. *The ACELT Journal*, *7*, 25-32. Manila, Philipines: Manila University.

White, R. & Arndt, V. (1996). *Process writing*. Harlow, Essex: Addison Wesley Longman Ltd.

Zamel, V. (1999). *A sourcebook for responding to student writing*, R. Straub (Ed.). Creskill, NJ: Hampton Press, Inc.

Acknowledgments

APPENDIX A
Profile Scale (Analytic Marking)
(Adapted from Liz Hamp-Lyons (1991))

COMMUNICATIVE ABILITY
ORGANIZATION
ARGUMENTATION
LINGUISTIC ACCURACY
LINGUISTIC APPROPRIACY

9. The writing displays an ability to communicate in a way which gives the reader full satisfaction.
The writing displays a completely logical organizational structure which enables the message to be followed effortlessly.
Relevant arguments are presented in an interesting way, with main ideas prominently and clearly stated, with completely effective supporting material; arguments are effectively related to the writer's experience or views.
The reader sees no error of vocabulary, spelling, punctuation or grammar.
There is an ability to manipulate the linguistic systems with complete appropriacy.

8. The writing displays an ability to communicate without causing the reader any difficulty.
The writing displays a logical organization structure which enables the message to be followed easily.
Relevant arguments are presented in interesting way with main ideas highlighted, effective supporting material and they are well related to the writer's own experience or views.
The reader sees no significant errors or vocabulary, spelling, punctuation or grammar.
There is an ability to manipulate the linguistic systems appropriately.

7. The writing displays an ability to communicate with few difficulties for the reader.
The writing displays good organizational structure which enables the message to be followed throughout.
Arguments are well presented with relevant supporting material and an attempt to relate them to the writer's experience or views.
The reader is aware of but not troubled by occasional minor errors of vocabulary, spelling, punctuation or grammar.
There are minor limitations to the ability to manipulate the linguistic systems appropriately which do not intrude on the reader.

6. The writing displays an ability to communicate although there is occasional strain for the reader.
The writing is organized well enough for the message to be followed throughout.
Arguments are presented but it may be difficult for the reader to distinguish main ideas from supporting material; main ideas may not be supported; their relevance may be dubious; arguments may not be related to the writer's experience or views.
The reader is aware of errors of vocabulary, spelling, punctuation or grammar, but these intrude only occasionally.
There is limited ability to manipulate the linguistic systems appropriately, but this intrudes only occasionally.

5. The writing displays a limited ability to communicate although there is often strain for the reader.
The writing lacks a clear organizational structure and the message is difficult to follow.
Arguments are inadequately presented and supported; they may be irrelevant; if the writer's experience or views

are presented their relevance may be difficult to see.
The reader finds the control of vocabulary, spelling, punctuation and grammar update.
There is inability to manipulate the linguistic systems appropriately, which causes severe strain for the reader.

4. The writing displays a limited ability to communicate which puts strain on the reader throughout.
The writing lacks a clear organizational structure and the message is difficult to follow.
Arguments are inadequately presented and supported; they may be irrelevant; if the writer's experience or views are presented their relevance may be difficult to see.
The reader finds control of vocabulary, spelling, or punctuation and grammar update.
There is inability to manipulate the linguistic systems appropriately, which causes severe strain for the reader.

3. The writing does not display an ability to communicate although meaning comes through spasmodically.
The writing has no discernible organizational structure and a message cannot be followed.
Some elements of information are present but the reader is not provided with an argument, or the argument is mainly irrelevant.
The reader is primarily aware of gross inadequacies of vocabulary, spelling, punctuation and grammar.
There is little or no cause of linguistic appropriacy, although there is evidence of sentence structure.

2. The writing displays no ability to communicate.
No organizational structure or message recognizable.
A meaning comes through occasionally but it is not relevant.
The reader sees no evidence of control of vocabulary, spelling, punctuation or grammar.
There is no sense of linguistics appropriacy.

1. A true non-writer who has not produced any assessable strings of English writing. An answer which is wholly or almost wholly copied from the input text or task is in this category.

0. Should only be used where a candidate did not attend or attempt this part of the test in any way.

APPENDIX B
Global Scale (Holistic Marking)
(Adapted from Liz Hamp-Lyons (1991))

BAND DESCRIPTORS

9   The writing displays an ability to communicate in a way which gives the reader full
satisfaction. It displays a completely logical organizational structure which enables the message to be followed
effortlessly. Relevant arguments are presented in an interesting way, with main idea prominently and clearly
stated, with completely effective supporting material; arguments are effectively related to the writer's experience
or views. There are no errors of vocabulary, spelling, punctuation or grammar and the writing shows an ability to
manipulate the linguistic system with complete accuracy.

8   The writing displays an ability to communicate without causing the reader any difficulties. It displays a
completely logical organizational structure which enables the message to be followed easily. Relevant arguments
are presented in an interesting way, with main ideas highlighted, effective supporting material and they are well
related to the writer's experience or views. There are no significant errors of vocabulary, spelling, punctuation or
grammar and the writing reveals an ability to manipulate the linguistic systems appropriately.

7   The writing displays an ability to communicate with few difficulties for the reader. It displays a
completely logical organizational structure which enables the message to be followed without much effort.
Arguments are well presented with relevant supporting material and an attempt to relate them to the writer's
experience or views. The reader is well aware of but not troubled by occasional minor errors of vocabulary,
spelling, punctuation or grammar, and/ or some limitations to the writer's ability to manipulate the linguistic
systems appropriately.

6   The writing displays an ability to communicate although there is an occasional strain for the reader. It is
organized well enough for the message to be followed throughout. Arguments are presented but it may be
difficult for the reader to distinguish main ideas from supporting material; main ideas may not be supported;
their relevance may be dubious; arguments may not be related to the writer's experience or views. The reader is
aware of errors of vocabulary, spelling, punctuation or grammar, and/ or limited ability to manipulate the
linguistic systems appropriately, but these intrude only occasionally.

5   The writing displays an ability to communicate although there is an occasional strain for the reader. It is
organized well enough for the message to be followed most of the time. Arguments are presented but may lack
relevance, clarity, consistency or support; they may not be related to the writer's experience or views. The reader
is aware of errors of vocabulary, spelling, punctuation or grammar, and/ or limited ability to manipulate the
linguistic systems appropriately.

4   The writing displays a limited ability to communicate which puts a strain on the reader throughout. It
lacks a clear organizational structure and the message is difficult to follow. Arguments are inadequately presented
and supported; they may be irrelevant; if the writer's experience or views are presented their relevance may be
difficult to see. The control of vocabulary, spelling, punctuation or grammar is inadequate, and the writer
displays an ability to manipulate the linguistic systems inappropriately, causing severe strain for the reader.

3   The writing does not display an ability to communicate although meaning comes through spasmodically.
The reader cannot find any organizational structure and cannot follow a message. Some elements of information

are present but the reader is not provided with an argument, or the argument is mainly irrelevant. The reader is primarily aware of gross inadequacies of vocabulary, spelling, punctuation and grammar; the writer seems to have no sense of linguistic appropriacy, although there is evidence of sentence structure.

2   The writing displays no ability to communicate. No organizational structure or message is recognizable. A meaning comes through occasionally but is not relevant. There is no evidence of vocabulary, spelling, punctuation or grammar, and no sense of linguistic appropriacy.

1   A true non-writer who has not produced any assessable strings of English writing. An answer which is wholly or almost copied form the input text or task in this category.

0   Should only be used where a candidate did not attend or attempt this part of the test in any way (i.e., did not submit an answer paper with his/ her name and candidate number written on).

APPENDIX C
Original Learner Questionnaire
(Ferris, 1995)

COMPOSITION SURVEY

1. How much of each composition do you read over again when your instructor returns it to you?
1st/ 2nd drafts
All of it ____   Most of it ____   Some of it____   None of it____
Final drafts
All of it ____   Most of it ____   Some of it____   None of it____

2. How many of your instructor's comments and corrections do you think about carefully?
1st/ 2nd drafts
All of it ____   Most of it ____   Some of it____   None of it____
Final drafts
All of it ____   Most of it ____   Some of it____   None of it____

3. How many of the comments and corrections involve:

1st/ 2nd drafts	A lot	Some	A little	None
Organization	____	____	____	____
Content/ Ideas	____	____	____	____
Grammar	____	____	____	____
Vocabulary	____	____	____	____
Mechanics	____	____	____	____

(e.g., punctuation, spelling)

Final drafts	A lot	Some	A little	None
Organization	____	____	____	____
Content/ Ideas	____	____	____	____
Grammar	____	____	____	____
Vocabulary	____	____	____	____
Mechanics	____	____	____	____

(e.g., punctuation, spelling)

4. If you pay attention to what your instructor write, how much attention do you pay to the comments and corrections involving:

1st/ 2nd drafts	A lot	Some	A little	None
Organization	____	____	____	____
Content/ Ideas	____	____	____	____
Grammar	____	____	____	____
Vocabulary	____	____	____	____
Mechanics	____	____	____	____

(e.g., punctuation, spelling)

Final drafts	A lot	Some	A little	None
Organization	____	____	____	____

Content/ Ideas   ____   ____   ____   ____
Grammar   ____   ____   ____   ____
Vocabulary   ____   ____   ____   ____
Mechanics   ____   ____   ____   ____
(e.g., punctuation, spelling)

5. Describe what you do after you read your instructor's comments and corrections (e.g., Do you look up the corrections in a grammar book? See a tutor? Rewrite your paper?
1st/ 2nd drafts _____
_____

Final draft _____
_____

6. Are there ever any comments or corrections that you do not understand? If so, can you give any examples?
_____

7. What do you do about those comments or corrections that you do not understand?
_____

8. Are there any of your instructor's comments positive? If so, can you give an example?
_____

9. Do you feel that your instructor's comments and corrections help you to improve your composition writing skills? Why or why not?
_____
_____

10. How would you rate yourself as a learner?
Excellent____   Good____   Fair____   Poor____

11. How would you rate your skills in writing compositions?
Excellent____   Good____   Fair____   Poor____

APPENDIX D
Learner Questionnaire
(Adapted from Ferris, 1995)

TO: English 102/ EC2 Students

   At this stage in your English 102/ EC2 class, you have already been exposed to a number of writing activities. These activities should have already given you insights about the writing practices commonly observed in a language composition class. To further study how you perceive and respond to them, please answer this questionnaire as honestly as you can. Please take note that the questions stated here mostly refer to the writing practices observed in your present language class except for some that require you to recall your previous writing experience. Thank you.

1. Are you aware of the process writing approach?
Yes_____   No_____
If yes,
a. how did you learn about it? _____
b. did you find it helpful when you first used it in writing?
Yes_____   No_____ Why or why not?_____
c. do you find it helpful now in your English 102/ EC2?
Yes_____   No_____ Why or why not?_____

2. How much of each composition do you read over again when your instructor returns it to you?
1st/ 2nd drafts
All of it ____   Most of it ____   Some of it____   None of it____
Final drafts
All of it ____   Most of it ____   Some of it____   None of it____

3. How many of your instructor's comments and corrections do you think about carefully?
1st/ 2nd drafts
All of it ____   Most of it ____   Some of it____   None of it____
Final drafts
All of it ____   Most of it ____   Some of it____   None of it____

4. How many of the corrections involve:

1st/ 2nd drafts	A lot	Some	A little	None
Organization	____	____	____	____
Content/ Ideas	____	____	____	____
Grammar	____	____	____	____
Vocabulary	____	____	____	____
Mechanics	____	____	____	____
(e.g., punctuation, spelling)				

Final drafts	A lot	Some	A little	None
Organization	____	____	____	____
Content/ Ideas	____	____	____	____
Grammar	____	____	____	____

	A lot	Some	A little	None
Vocabulary	____	____	____	____
Mechanics	____	____	____	____

(e.g., punctuation, spelling)

5. If you pay attention to what your instructor writes, how much attention do you pay to the comments and corrections involving:

1st/ 2nd drafts	A lot	Some	A little	None
Organization	____	____	____	____
Content/ Ideas	____	____	____	____
Grammar	____	____	____	____
Vocabulary	____	____	____	____
Mechanics	____	____	____	____

(e.g., punctuation, spelling)

Final drafts	A lot	Some	A little	None
Organization	____	____	____	____
Content/ Ideas	____	____	____	____
Grammar	____	____	____	____
Vocabulary	____	____	____	____
Mechanics	____	____	____	____

(e.g., punctuation, spelling)

6. Describe what you do after you read your instructor's comments and corrections (e.g., Do you look up the corrections in a grammar book? See a tutor? Rewrite your paper?
1st/ 2nd drafts _____
_____

Final draft _____
_____

7. Are there ever any comments or corrections that you do not understand? If so, can you give any examples?
_____

8. What do you do about those comments or corrections that you do not understand?
_____

9. Are there any of your instructor's comments positive? If so, can you give an example?
_____

10. Do you think that your instructor's comments and corrections help you to improve your composition writing skills? Why or why not?
_____
_____

11. Do you feel that your instructor's comments and corrections help you discover and/ or improve your skills in revising your paper? Why or why not?

_____

_____

12. Which form of assessment do your prefer?
One with overall numerical mark for the composition _____
One with detailed comments and a numerical mark   _____

13. Generally, how do you find your teacher's feedback on your composition?
Positive_____   Neutral_____   Offensive_____

14. How often do you receive such kind of feedback on one writing assignment?
Never_____   Rarely_____   Seldom_____   Occasionally_____   Always___

15. How would you rate yourself as a learner?
Excellent_____   Good_____   Fair_____   Poor_____

16. How would rate your skills in writing compositions?
Excellent_____   Good_____   Fair_____   Poor_____

17. Generally speaking, I find teacher's comments at the end of my paper
Helpful_____   Useless_____

18. Which of the following drafts do you get feedback on by way of conferencing?
1st/ 2nd drafts_____   Final draft_____

19. Are there any comments or corrections that you do not understand during conferencing? If so, can you give any example?_____

20. Do you finish conferencing with your teacher without understanding some of the comments or corrections?
Yes_____   No_____ If yes, can you given an example?_____

21. Do you think conferencing helps you clarify your thoughts/ ideas with your teacher?
Yes_____   No_____ If no, why not?_____

22. Do you think conferencing helps your teacher clarify his/her thoughts/ ideas with you?
Yes_____   No_____ If no, why not?_____

23. Do you think conferencing helps you discover and/ or improve your skills in revising your paper?
Why or why not?_____

24. Generally speaking, I find conferencing
Helpful_____
Useless_____

Investigating the Influence of Transcribing, Reporting, and Task Repetition on In-Class Spoken Task Performances

Jason Moser
(jason.moser@osaka-shoin.ac.jp)
Osaka Shoin Women's University, Japan

Abstract

This study is a classroom-based, quantitative study into the influence of transcribing, reporting, and task repetition on in class student oral task performances. The study investigates two questions. First, do students in an intact classroom improve task performances when they repeat the same task in subsequent performances? In a previous exploratory study by Moser (2008), students did not take advantage of task repetition opportunities to improve a repeat task performance. It was concluded that the reason for this was that amongst many students, there was a lack of perceived pedagogical rationale for task repetition. This current study thus investigates whether a more transparent pedagogical focus realized through a transcribing phase or a reporting phase prior to a repeat task performance result in improved subsequent task performances. Related to this, and the second question of this study, is does the more intensive transcription work result in improved task performances than the reporting work? The results of the study reveal no significant difference between transcribing or reporting on subsequent task performances; however, there were significant results for a task repetition effect on task performances. The classroom implications of these findings will be discussed.

Key words: task-based learning, task repetition, transcription, reformulation, AS-units, working memory

Introduction

Task repetition research over the past decade has demonstrated that task repetition (TR) is potentially a useful pedagogical option for teachers trying to ease the processing pressures students face during task work (see Ahmadian and Tavakoli 2011, Bygate 1996, 2001, Bygate and Samuda 2005, Gass et al. 1999, Hawkes 2012, Lynch and Maclean 2001, and Pinter 2005). Since Bygate's first study, TR has become a mainstream concept in ELT, and is widely recommended in teacher support texts (see Thornbury 2005, Willis and Willis 2007). As a pedagogical option, it involves repeating the same or similar task at intervals. TR researchers have argued that it benefits students in two basic ways. First, it facilitates easy retrieval of prior content, and thus allows students wider access to input that can be used to expand production. Second, students, through freed-up processing capacity, are better able to monitor or attend to language form in subsequent performances. The potential for students to focus on both meaning and form simultaneously is why Bygate and Samuda (2005, p. 43) refer to TR as "integrative planning." For this reason, they believe that it is more effective than either online planning or strategic planning.

TR studies, like most task-based studies, have until very recently been done mostly in controlled settings, and have been predominately cognitive-based studies. Samuda and Bygate (2008) rightly note that there is a need to validate the effectiveness of task-based learning in actual classroom settings. Specifically, with TR studies there has been an over-reliance on narrative tasks despite the fact that most oral competency-building classes (see

Hawkes 2012 as an example) use a variety tasks with the majority of these being dialogical rather than monological.

An exploratory classroom study by this author (Moser 2008), which was the impetus for this current study, took up Bygate's (1996) call to investigate what students actually do in the classroom with TR when unguided and unprompted by the teacher. To use William's (2005, p. 679) term, the "locus of responsibility" was on the students to take advantage of the opportunities afforded by task repetition. In this first exploratory study, students in large intact classrooms (35+) did the same thematic conversation-like pair work task three times in one class with a different partner each time. The interval between tasks was immediate. When the data was reviewed, the overall trend demonstrated that the majority of students' repeat task performances in terms of productivity and overall quality either did not improve or deteriorated. Ohta's (2001, p. 256) classroom-based study also found that with TR it was common for students who were repeating a task to use more minimal and elliptical language than in the initial task. Very few students in my study engaged in what Bygate (2006, p. 168) calls "constructive repetition." I concluded that most of the students neither recognized nor acted on the pedagogical purpose of immediate TR. Two important findings emerged. First, using tasks and task options do not automatically result in productive learning. As Samuda and Bygate (2008, p. 35) note, there has been a tendency in task-based learning literature to assume that tasks can facilitate learning processes "by magic." I concluded that in addition to having a communicative goal for a task, there also needs to be a clear form-focused goal as well, and that the latter in some classroom contexts like mine is probably more important than the former for ensuring student engagement. Second, the study indirectly highlighted another point also recognized by Hawkes (2012), and that is TR studies to date have not provided any detailed guidelines on how teachers and students go about improving repeat task performances.

This current study then became an action research study intending to investigate if a more transparent pedagogical focus accompanying TR would result in improved student oral task performances. In order to make the pedagogical purpose of TR more transparent it was decided to add between the task repetitions analytical language learning activities, in particular student reporting or student self-transcribing. The reporting strategy is based on Willis's (1996) task cycle where students report to the class and teacher on the contents of their task after completing it. According to Willis, reporting pushes the students to use their best language, and as a result focus on language form. The idea of student transcribing of an oral task performance is from Lynch (2001,2007) who hypothesized that making a student's language from his/her oral task performance visually salient and thus available for analysis and reformulation would increase the probability of the student reprocessing highlighted forms in a subsequent task performance. In his latter controlled study he found that the active transcribers, despite being the less proficient group, achieved a higher level of accuracy in a subsequent task performance on items that had been focused on in the transcription process. Lynch (2007) concluded that transcribing could be an effective classroom tool for getting students to improve their accuracy in task performances. Other studies (Mennim 2007, Stillwell et al. 2010, Stones 2012) have also found similar positive findings for using student transcription to improve subsequent task performances.

Research Questions

The first question for this study was: Do transcription and reporting between iterations within a TR sequence facilitate student task engagement and improved language performance? As a supplement to this first question, like Bygate (2001), a repeat performance after a longer gap in instruction was also included in the research design, in this case incorporating data from student production three to four weeks after the second lesson.

The second question considered whether there would be any difference in results between the reporting and transcription groups. The positive findings of Lynch's studies support the hypothesis that the self-transcription group will outperform the reporting group in subsequent task performances. The differences between the two treatments are as follows:

Reporting: recalled, teacher-initiated feedback, teacher feedback limited and general (whole class).
Transcribing: visually available, student-initiated feedback, teacher feedback detailed and specific.
Thus, the second research question was: Does the transcription group outperform the reporting group in repeated task repetitions after pedagogical interventions?

Method

Participants

The study involved two intact classes of 20 students from a less academically competitive women's university in Japan. The students were second year non-English majors from two different departments. The actual intervention covered the full 90 minutes of 12 of the 15 classes in the first semester. Based on a local placement test (EIKEN) all of the students in this study could be categorized as low beginners. The very top scores on the placement test were roughly equivalent to 3.5 on the IELTS test.

Design

In an intact classroom study, there were a number of constraints that influenced the design. As Rossiter (2001, p. 36) writes, often the most challenging part of doing classroom-based research involves the use of intact classrooms. In particular, she explains that proficiency can vary widely within a class as well as between classes. Specifically, with this study the two classes available were mostly incomparable as one class, as indicated by the placement test, had a larger group of students with a relatively higher level of proficiency. This difference was because one department was more competitive to enter than the other. Scheduling and university policy made it impossible to create two ideal classes to compare. The two classes were also chosen because the students in both classes had enough English ability to make the research proposal feasible. Most of the other classes potentially available were just too low to conduct the intended research.

In order to overcome the difference in the two intact classes, I decided to create, based on the placement test scores, two artificial groups from the 40 students. A subsequent paired-samples t-test indicated no difference between the two new groupings. What this meant then in terms of design and actual classroom teaching is that in each class there would be two groups of 10 students doing a different treatment at the same time. The reality of having to do two treatments at the same time in classes created another issue in that it was impossible to have one group always do the more intensive and demanding transcription. Students doing the more intensive transcribing would have eventually protested against the extra work, which would have resulted in having to change or abandon the study. For this reason each group completed the transcription treatment twice and reporting treatment twice in an alternating pattern. Although all the students contributed to the data for each treatment, it was treated statistically as if the students contributed to only one of the treatments. Finally, as part of the study the students were required to keep a learning journal as well as complete a survey at the end of the course.

Tasks

Throughout this study the term task session will be used to refer to the four open-ended topic interview tasks; pets and animals, shopping and fashion, health and exercise, and dating and marriage. In each task session there are four task performances. They are referred to as a first task performance, second task performance and so forth. None of these topics were used in the students' other English class. Students did the task sessions in the same order. Due to absences not every student did all four task sessions. The basic task for this study involved pairs taking turns being the interviewer and interviewee with a set of provided questions. Students were told that the interviewee was the focus of the task, and it was primarily the interviewee's responsibility to do the task well. Each student as an interviewee was recorded for six minutes (+/- 30 seconds).

Lesson Sequences for Transcribing and Reporting

The two lesson treatments on one topic spanned two classes of two weeks, as outlined below.

Transcribing Group	Reporting Group
First Class	**First Class**
1. Unguided pre-task planning (15-20) minutes.	1. Unguided pre-task planning (15-20) minutes.
2. Do task with partner.	2. Do task with partner.
3. Transcribe performance.	3. Recall and write report as dialogue.
4. Self-correct transcript.	4. Read report with teacher feedback for group.
5. Edit and reformulate transcript.	5. Edit and reformulate recalled dialogue.
6. Answer self-reflection questions and hand in notebook.	6. Answer self-reflection questions and hand in notebook.

Transcribing Group	Reporting Group
Second Class	**Second Class**
1. Teacher returns notebooks.	1. Teacher returns notebooks.
2. Teacher-directed whole class feedback session Positive and negative evidence of language used plus new native speaker input introduced.	2. Teacher-directed whole class feedback session Positive and negative evidence of language used plus new native speaker input introduced.
3. Students review transcript and teacher feedback, finish editing and reformulating their transcript.	3. Students review dialogue and finish editing and reformulating.
4. Students do a rehearsal of same task with first partner.	4. Students do a rehearsal of same task with first partner.
5. Students are assigned a new partner and do the same task again.	5. Students are assigned a new partner and do the same task again.
6. Students answer self-reflection questions in their journal and are assigned next week's homework.	6. Students answer self-reflection questions in their journal and are assigned next week's homework.
7. Hand in notebooks Teacher listens to transcribed second and third performances and gives limited but updated feedback.	7. Hand in notebooks. No feedback provided for second or third performances.

Transcribing Group	Reporting Group
Three/Four weeks later	**Three/Four weeks later**
A review session is held and the students do the task again with a new partner from their group.	A review session is held and the students do the task again with a new partner from their group.

Identifying Reading

The classroom realities of this study made it impossible to completely prohibit students from referencing their notes. During the trialing stage, students strongly insisted that they have access to their notes. This "borrowing" (Prabhu, 1987) is "necessary for maintaining task-based activity" (p. 61), since it probably has some "direct value"

for acquisition. While students were discouraged from using their notes, they were allowed to reference them in every task performance except the fourth task performance, which was designated a review task. In all four task performances, the interviewer had access to the task questions. To make it difficult for the students to use notes, they were required at the start of each task to close their notebooks, and place them off to the side of their desk.

In order to accurately codify the data it was still necessary to account for student reading and exclude it. To do this, two measures were introduced after the trialling period. First, the students after each task performance were given a corresponding coloured pen to underline parts of their notes, reports, and transcripts, they read or clearly looked at. To facilitate honesty, students were told that in no way would the reading of notes affect their final grade, unless they made no effort to speak without them. There were only a few habitual readers who were also very honest about their reading. For many students, not using notes became a primary goal for them in the course. Second, the teacher kept a record of students that were reading or overusing their notes. It was also relatively easy with this level of students to identify reading when reviewing their recorded task performances.   When a student did read, her speech was usually monotone. This was usually preceded by nonverbal cues like the rustling of paper accompanied by long periods of silence or L1 use.

Measures
In this study students' oral performances were divided into AS-units (Analysis of Speech Unit) (Foster et al. 2000). An AS-unit is a main clause and any attached subordinate clauses or sub-clausal units. While the AS-unit is primarily a syntactical unit, Foster et al. argue that for the AS-unit to be valid, it must reflect the psycholinguistic processes of what a learner can do in a 'single unit' of micro-planning.   As part of this cognitive unit researchers must also take into consideration intonational and hesitational markers.   The second level in Foster et al.'s transcription recommendations was used including their suggestions for dealing with dysfluent language. Single-word or two-word units in AS-unit totals were not counted. Additionally, three-word AS-units functioning as simple responses (*Yes I can, Yes I do, Yes I am*) were generally excluded unless the student's performance was very weak.

Complexity Accuracy Fluency (CAF)
Norris and Ortega (2009: 575) warn researchers CAF it is not "some kind of universal construct" that can be blanketed across all contexts without adaptation and remind us that researchers must consider their context and what CAF is supposed to reflect. On this point one of the challenges of using CAF is to adapt the construct to the context in which the research is being conducted to accurately quantify task performances. The CAF measures used in this study were selected or adapted by first analyzing the students' oral task performances. What emerged was the realization that because of the very basic features of the students' production that general measures had to be used over more specific or finer measures to capture any potential variance across task performances. Tavakoli and Skehan (2005: 256) note that general measures while more "blunt" have been shown to capture variance in task performance better than specific measures.

Complexity
For complexity the following measures were used:
- Total words per task performance
- Words per AS-unit
- Number of AS-units seven words and over
- Total words of AS-units seven words and over
- Largest single AS-unit.

The measure total words per task performance reflects student language productivity. Simply, the more words a student produces the more productive he/she is, which Dörnyei and Kormos (2000) consider a reflection of student task engagement. The measure words per AS-unit was used to measure how many words a student could

on average produce in a speech unit. This seemed suitable considering that most of the students relied on rule-based processing, and therefore produced a lot of language word by word.

The two complexity measures number of AS-units seven words and over and total word count of AS-units seven words and over were measures that attempted to identify how much of a student's task performance was at the upper boundaries of their interlanguage. The idea is from Skehan and Foster's (2005: 198) accuracy benchmark, which involved discovering a 'cut-off point" in terms of length of clause where students cannot produce correct clauses. In a previous study (Moser 2008), I explored the idea of a complexity benchmark and analysed at what word count an AS-unit started to reflect challenging production for the students in that study, determining the border was at six- and seven-word AS-units. At the six- and seven-word mark there was an increase in error-filled AS-units and an increase in conjunctions within the AS-units. The same benchmark of seven words was used for this study because the students here were noticeably less proficient than the students in the previous study. Benchmark in this study then refers to a starting point where it was determined students' AS-units reflect complex language use. The largest single AS-unit was used as well because one AS-unit in a performance could comprise up to 30 words and would often represent the most complex part of a task performance.

Accuracy

For accuracy four measures were used:
- Number of error-free AS-units (includes performances errors)
- Total number of error-free AS-units of seven words and over
- Total number of words from error-free AS-units of seven words and over
- Number of reformulations (both grammatical and lexical).

In this study, errors refer to errors in syntax, word order, morphology, and lexical choice using native-like accuracy as a comparison model. Performance errors were included and refer to errors that clearly did not appear to be a slip of tongue. Because of the beginner level of the students, in this study I chose error-free AS-units rather than Bygate's (2001) errors per unit. During the codifying of the data, it became clear that when students were being productive and engaging in risking taking by producing longer AS-units, these AS-units almost always contained multiple errors. Based on a preliminary analysis of the data, I concluded that if I used this measure, incidence of error per AS-unit would increase with more productive task performances. The next two accuracy measures based on the established benchmark measured how much of the best part of a student's production was accurate. The measure number of reformulations per task performance indicates an orientation on the part of the student to speak accurately, with reformulations defined as changes to syntax, word order, and morphology as well as replacements or lexical corrections/substitutions. This measure would provide information on whether students could attend to language form while their attention was primarily focused on meaning.

Fluency

The following measurements were used for each sub-dimension of fluency:
- Speed: Mean length of run
- Repair: Total number of dysfluent words (false starts, repetition of words phrases, reformulations, replacements)
- Breakdown: Number of filled and unfilled pauses (only pausing that was over five seconds and occurred within AS-unit boundaries was counted).

In my study, length of run is a stream of speech measured in words sandwiched between clear pauses, dysfluent language, and L1 (Freed 2000) and not syllables per run as that measure was not appropriate for the data gathered, particularly since the students seemed to have trouble isolating individual English phonemes and thus

unnecessarily created a number of extra syllables. I was more interested in how many words students could produce in one spurt rather than the speed of that spurt. In addition, only length of run in AS-units of seven words and over with at least two clauses or subclauses were counted. The final two measures for fluency were chosen because both appeared in abundance in the data.

Analytical Procedure

This study used a three-way ANOVA with two factors varied between-subjects (transcribing versus reporting, and four task sessions) and one within-subject factor (four performances of each of the task sessions). Thus, there were four different task sessions (four themes/topics) and each task session was performed four times (four task performances). The three independent variables in the study were the treatments, the four task sessions, and task repetition. Because of space constraints and lack of impact on the study, the task session results are not reported. The data contains only complete task sessions. The 40 students who were involved in the study produced a sample size of n=119.

Table 1
Participants Across Treatments and Task Sessions

Task session	Reporting students	Transcribing students	Total
1	13	14	27
2	12	18	30
3	15	18	33
4	14	15	29
			119

Results

Transcribing versus Reporting Results

As Table 2 shows, none of the CAF measures reached significance, indicating neither transcribing nor reporting significantly influenced any of the measures used. In relation to the research question: Does the transcription group outperform the reporting group in repeated task repetitions after pedagogical interventions?, the results are straightforward in that students who did the transcribing treatment did not produce more improved subsequent task performances than the reporting treatment. As will be explained in the discussion section, students in their survey feedback and journals reported a lot difficulty with reformulation work and trying to incorporate it in their subsequent task performances.

Task repetition results

For the within subjects tests there were significant effects across task repetitions for all CAF measures except number of pauses.

Table 2
Between-subjects Results

Measure	source	ss	df	ms	f	sig
Total words per task performance	treat	1800.338	1	1800.36	.799	.373
	task	7078.088	3	2359.363	1.048	.374
	treat+task	6694.528	3	2231.509	.991	.400
Words per AS-unit	treat	.084	1	.084	.014	.906
	task	11.060	3	3.687	.611	.609
	treat+task	3.105	3	1.035	.171	.915
Number of AS-units seven words and over	treat	5.393	1	5.393	.427	.515
	task	17.279	3	5.760	.456	.713
	treat+task	21.268	3	7.089	.562	.641
Total words of AS-units seven words and over	treat	538.625	1	538.625	.255	.614
	task	2963.660	3	987.887	.468	.705
	treat+task	2258.374	3	752.791	.357	.784
Largest single AS-unit	treat	.029	1	.029	.000	.983
	task	62.219	3	20.740	.314	.815
	treat+task	39.811	3	13.270	.201	.896
Number of error-free AS-units	treat	9.640	1	9.640	.558	.456
	task	369.633	3	123.211	7.137	.000
	treat+task	114.115	3	38.038	2.203	.092
Number of error-free AS-units seven words and over	treat	.844	1	.844	.222	.638
	task	71.136	3	23.712	6.248	.001
	treat+task	6.230	3	2.077	.547	.651
Total number of words of error-free AS-units seven words and over	treat	242.014	1	242.014	.636	.427
	task	7740.073	3	2580.024	6.785	.000
	treat+task	498.727	3	166.242	.437	.727
Number of reformulations	treat	.238	1	.238	.068	.795
	task	22.800	3	7.600	2.171	.095
	treat+task	14.167	3	4.722	1.349	.262
Mean length of run	treat	1.210	1	1.210	.345	.558
	task	42.295	3	14.098	4.014	.009
	treat+task	1.474	3	.491	.140	.936
Number. of dysfluent words	treat	420.784	1	420.784	1.888	.172
	task	702.683	3	234.228	1.051	.373
	treat+task	486.154	3	162.051	.727	.538
Number of pauses	treat	10.245	1	10.245	2.628	.108
	task	44.148	3	14.716	3.775	.013
	treat+task	23.229	3	7.743	1.986	.120

Complexity

To confirm significance for the repeated measures design of this study, Mauchly's test of sphericity was first checked for a measure to confirm equal variance across conditions and subjects. For the complexity measures, number of words per task performance (p = .225), and words per AS-unit (p = .855) the Mauchly's test of sphericity was non-significant. The sphericity assumed tests were significant for both measures (F=27.252, p ≤ . 001 and F=9.554, p ≤ .001). Table 3 contains the estimated marginal means for all the complexity measures.

Table 3
Descriptive Means for Complexity

Measure	Task session	Mean	Std. error	95% confidence interval	
				Lower bound	Upper bound
Total words per task performance	1	75.322	2.544	70.281	80.364
	2	85.356	2.984	79.444	91.269
	3	99.491	3.015	93.517	105.464
	4	94.695	2.715	89.315	100.075
Words per AS-unit	1	6.418	.135	6.150	6.686
	2	6.957	.148	6.663	7.250
	3	7.178	.155	6.871	7.484
	4	6.830	.140	6.553	7.107
Number of AS-units seven words and over	1	3.877	.186	3.508	4.246
	2	4.900	.219	4.467	5.334
	3	5.994	.253	5.492	6.496
	4	5.511	.239	5.038	5.984
Total words of AS-units seven words and over	1	38.594	2.256	34.123	43.065
	2	50.841	2.785	45.323	56.358
	3	62.861	3.062	56.794	68.928
	4	56.175	2.799	50.628	61.722
Largest single AS-unit	1	12.605	.452	11.709	13.500
	2	14.059	.470	13.127	14.990
	3	15.746	.551	14.655	16.837
	4	14.623	.537	13.560	15.687

The confidence intervals for total words per task performance revealed a number of significant pairwise comparisons. The upper bound of the first task performance (80.364) was significantly lower than the lower bounds of task performance three (93.517) and task performance four (89.315). In addition the upper bound of the second task performance (91.269) was less than the lower bound of third task performance (93.517). These results are further supported by the polynomial trends result which confirms the means of the conditions and whether they fit a trend. The results for total words per task performance revealed a significant linear trend (F = 66.741, p ≤ .001) of a straight line rising across the task repetitions. For the next measure words per AS-unit,

there was a significant pairwise comparison with the upper bound of the first task performance (6.686) less than the lower bound of the third task performance (6.871). There was also a significant linear effect (F = 10.520, p = .002).

For the final three complexity measures (number of AS-units seven words and over, total words of AS-units seven words and over, and largest single AS-unit), there were significant main effects for task repetition. Mauchly's test of sphericity for all three measures was significant. For number of AS-units seven words and over, Mauchly's test of sphericity was on the edge of significance (p = .050). The Huynh-Feldt correction test was thus referred to and indicated significance (F = 26.117, p ≤ .001). The pairwise comparisons for number of AS-units seven words and over revealed a number of significant differences. The upper bound of the first task (4.246) was significantly lower than the lower bounds of task performances two (4.467), three (5.492) and four (5.038). In addition the upper bound of the second performance (5.334) was significantly less that the lower bound of the third performance (5.492). The polynomial trends indicated a significant linear effect (F = 54.395, p ≤ .001). For the total words of AS-units seven words and over, as mentioned, Mauchly's test of sphericity was significant (p = .045) and the Huynh-Feldt correction test was significant (F = 26.357, p ≤ .001). The pairwise comparisons revealed significant differences between the task performances. Task performance one's upper bound (43.065) was significantly lower than the lower bound of task performance two (45.332), task performance three (56.794) and task performance four (50.528). Task performance two's upper bound (56.358) was also significantly lower than task performance three's lower bound (56.794). Lastly, there was also a significant linear effect (F = 52.541, p ≤ .001). The final measure, largest single AS-unit per performance, had a significant Mauchly's test of sphericity (p =.033). The Huynh-Feldt correction test was significant (F = 11.352, p ≤ .001). Table 3 shows pairwise comparisons that revealed that the first task performance's upper bound (13.500) was significantly lower than both the third task performance (14.655) and the fourth task performance (13.560). The polynomial trends test also indicated a significant linear effect (F = 18.382, p ≤ .001).

In summary, for complexity the results showed that in relation to the third performance the first task performance was significantly less complex on all complexity measures. For three measures the first task performance was significantly less complex than all three subsequent tasks performances. The third task performance besides being significantly more complex than task performance one on all measures was significantly more complex than task performance two on three measures. Overall, for complexity, the data suggest that the first task performance was the weakest relative to the subsequent task performances, and the third task performance the most complex relative to the first and second task performances.

Accuracy

Results across task repetitions were significant for all accuracy measures. For number of error-free AS-units, Mauchly's test of sphericity was non-significant (p = .120). The ANOVA test was significant (F = 4.977, p = .002). The polynomial trends showed a significant linear effect (F = 9.078, p = .003).

As shown in Table 4, the mean scores follow the main pattern observed throughout the CAF measures. The first performance had the least number of error-free AS-units and the third the most. The pairwise data showed that there was a significant difference between task performance one and task performance three with the higher bound of the first task performance (6.168) lower than the lower bound of the third task performance (6.337). For the next two measures (number of error-free AS-units seven words and over and number of words from error-free AS-units seven words and over), the Mauchly's test of sphericity was non-significant for each (p = .107 and .106). Both ANOVA tests were significant (F = 16.502, p ≤ .001 and F = 15.014, p ≤ .001). Polynomial trends results indicated significant linear trends for both measures (21.078, p ≤ .001 and F = 18.789, p ≤ .001).

For total number of error-free AS-units seven words and over, the pairwise comparisons showed that the lower bound of third task performance was significantly higher (2.123) than the upper bounds of task performance one (1.467), task performance two (2.033) and task performance four (2.085). In addition, the

fourth task performance's lower bound (1.538) was significantly higher than the upper bound of the first task performance (1.467). For total number of words of AS-units seven words and over, the descriptive means revealed significant pairwise differences. he upper bound of the first task performance (13.423) was significantly lower than task performance two (14.008), task performance three (19.563), and task performance four (14.030). In addition, the third task performance's lower bound was significantly higher than the upper bound of the fourth task performance (19.218). The final accuracy measure number of reformulations, which reflects an orientation towards being accurate, did not have a significant result for Mauchly's Sphericity test (p = .193). The accompanying ANOVA test was significant (F = 4.899, p = .002,) and there was a significant linear effect (F = 12.395, p = .001). There was one significant pairwise comparison. The upper bound of task performance one (1.370) was significantly lower than the lower bound of task performance four (1.522).

Table 4
Descriptive Means for Accuracy

Measure	Task session	Mean	Std. error	95% confidence interval	
				Lower bound	Upper bound
Number of error-free AS-units	1	5.628	.272	5.088	6.168
	2	6.076	.302	5.478	6.674
	3	6.886	.277	6.337	7.435
	4	6.352	.270	5.817	6.886
Number of error-free AS-units seven words and over	1	1.275	.097	1.082	1.467
	2	1.746	.145	1.459	2.033
	3	2.429	.154	2.123	2.734
	4	1.811	.138	1.538	2.085
Total number of words of error-free AS-units seven words and over	1	11.498	.971	9.574	13.423
	2	17.000	1.510	14.008	19.992
	3	22.700	1.583	19.563	25.836
	4	16.624	1.309	14.030	19.218
Number of reformulations	1	1.116	.128	.862	1.370
	2	1.485	.134	1.219	1.751
	3	1.569	.158	1.256	1.882
	4	1.813	.147	1.522	2.103

In summary, for accuracy the results revealed that for every measure the first task performance was significantly less accurate than at least one subsequent task performance. With the measure number of words from error-free AS-units seven words and over, the first task performance was significantly less accurate then all three subsequent task performances. The third task performance with the exception of the number of

reformulations was significantly more accurate than task performance one on three measures and in addition significantly more accurate than all performances for the measure number of AS-units seven words and over. It was also significantly more accurate than the fourth performance on number of words of AS-units seven words and over. The fourth performance was more accurate than the first performance on two measures. As with complexity the results suggest that task performance one for accuracy measures was the least accurate while the third task performance was generally the most accurate.

Fluency
Results across task repetitions were significant for length of run and number of dysfluent words. For number of pauses, there was no significant effect for task repetition.

Table 5
Descriptive Means for Fluency

Measure	Task session	Mean	Std. error	95% confidence interval	
				Lower bound	Upper bound
Mean length of run	1	5.628	.098	5.760	6.168
	2	6.442	.115	6.215	6.669
	3	6.631	.119	6.395	6.868
	4	6.334	.111	6.115	6.554
Number of dysfluent words	1	9.838	.807	8.239	11.436
	2	13.585	.913	11.775	15.394
	3	14.149	.953	12.261	16.038
	4	15.268	.978	13.329	17.206

As indicated in Table 5, Mauchly's test of sphericity for length of run was non-significant ($p = .636$). The follow-up ANOVA test was significant ($F = 12.761$, $p \leq .001$). There was also a significant linear effect ($F = 16.149$, $p \leq .001$,). The pairwise comparisons revealed significant effects. The upper bound of task performance one (6.148) was significantly lower than the lower bound of task performance two (6.215) and the lower bound of task performance three (6.395). For number of dysfluent words per task performance Mauchly's test of sphericity was non-significant ($p = .264$) and the ANOVA test was significant ($F = 11.523$, $p \leq .001$). There was a strong linear effect ($F = 23.288$, $p \leq .001$,), which means there was a substantial increase from the first performance to the last performance. In this study with the level of students the more dysfluent language there was the more fluent the task performance was. The pairwise comparisons revealed that the upper bound of the first task performance was significantly lower than the lower bound of task performance two (11.775), task performance three (12.261) and task performance four (13.329).

In summary, the results for fluency reveal that on the two measures length of run and number of dysfluent words the first task performance was the less fluent. For length of run it was significantly less fluent than task performance two and three. For number of dysfluent words, the first task performance had significantly fewer dysfluent words than each of the three other task performances, which, as mentioned, reflects less productive language production.

The task repetition results apply to the first question of this study: Do transcription and reporting between iterations within a task repetition sequence facilitate student task engagement and improved language

performance? Taking into consideration the findings from Moser (2008), which was the rationale for this study, there appears to be some support for the idea that transparent pedagogical interventions or learning focuses resulted in improved subsequent task performances. At a minimum, the within-subjects results suggest that task repetition in a classroom setting results in improved subsequent task performances.

Discussion

Transcription versus Reporting Discussion

There are a number of interrelated classroom variables that should explain why the transcribers did not outperform the reporters on any CAF measure in subsequent task performances.

First, minor factors were in the features and outcomes of the error correction and reformulation processes. Most of the errors in this study involved tense, subject verb agreement, and grammatical words. In any one performance there would be an abundance of these errors, and many of these, in particular grammatical words, are often resistant to error correction. Despite the visual saliency of these errors in the students' transcripts, in hindsight an extensive 'correct everything' approach appeared to be ill suited for these types of errors. The error correction approach combined with reformulation work also left the final transcripts looking very messy and confusing to the degree that students often appeared to quickly glance over their final transcripts rather than analyze them.

A more important set of factors that contributed to the non-results for transcription versus reporting were the students' low proficiency level combined with the challenges of a first task performance. This combination meant that first task performance resulted in minimal language production to the point where there was not a lot of language to correct or work on for the teacher and student. Besides self-correcting, students were asked to reformulate their transcripts. Many students were already working at the limits of their interlanguage often could not add more detail on their own at the reformulation stage. When they did add detail because of their low proficiency, this mostly meant chaining clauses together to extend their utterances rather than trying to make their utterances more precise or inserting more meaning into them. Analysis of student transcripts showed that very few students applied reformulation work nor teacher provided input in subsequent performances unless they memorized it, and produced it at the start of a task performance.

The fact that students did not reattempt corrected errors or incorporate reformulated language in their subsequent task performance suggests that probably a working memory constraint (see Dai 2013) was the most substantial factor in the transcribing versus reporting results. For students to incorporate corrected units, apply points of grammar or reformulated work into subsequent performances, it required them to retain substantial amounts of language in their working memory and then employ it appropriately during a real-time task performance. The working memory constraint appears to have negated any benefits that a more intensive and visually salient transcription treatment could provide. While not explicitly reported as a working memory constraint, the students in Stillwell et al.'s (2010) transcription study also appeared to have had working memory issues, as their findings report that some students were unsuccessful at incorporating changes made at the reformulation stage in a repeat oral task performance.

Task Repetition

First Performance and Second Performance

As the results indicated the first performance, despite being supported by pre-task planning, was for the most part the weakest of the task performances. A lot of the production appeared to be word by word processing. Dysfluent language did not emerge from production breakdowns or repair fluency, but rather emerged as students struggled to start speaking in what could be 'start up dysfluency'. The first performances clearly reaffirmed that most students in the study had a very low level of language proficiency. For the second performance, the results show that on a number of measures it was significantly weaker than the third performance. Also in comparison to the third and fourth task performances, it had less significant pairwise

comparisons in relation to the first task performance. Overall, the data suggests that it was probably the second weakest of the four task performances. The reason for this is fairly straightforward. In the original trialling of the study a repeat task performance was supposed to occur only once before a final task performance three to four weeks later. However, it became apparent during initial trailing that in the follow-up task performance students were not focusing on employing corrected language or reformulations. In order to facilitate student reworking of prior work into a subsequent task performance, and furthermore confirm whether students could in fact do it, another task performance (second task performance) was added where students had to work with the same partner and deliberately try to improve on their first performance. As discussed earlier, it appears that working memory constraints prevented this. In addition, the students similar to the students in the Ohta (2001) study appeared to get bored from the rehearsal-like nature of the second performance, and as a result started to use minimal and elliptical language.

Third Performance
The results indicated that overall the third performance could be considered the strongest especially in comparison to the first task performance. An important variable appears to have been the proximity of the second performance to the third. This immediate proximity made verbatim-like language carry-over from the second performance to the third performance more frequent. Task familiarity also made it easier for students to produce language on a new strand of the task topic untraceable to previous performances. This language represented the majority of the increased language production for most students in the third performance. Another key characteristic of third performance was the addition of new propositional content on an earlier centre of interest. The extracts (see Appendix for transcript conventions) below are rare example of a centre of interest being repeated and expanded on from the first to the third task performance.

Task Performance 1
1P: How often do you exercise?
2S: /{I}[I play tennis]{5.70}{two}[two times a week]/
3P: Only tennis?
4S: /Yeah/
5P: Then what sports or exercise you used to do?

In the second performance below the student repeats from the first performance that she plays tennis twice a week, adding new information at 17S. This new information is notably facilitated by her partner's question at 16P.

Task Performance 2
14P: Do you exercise?
15S: /{I} I play tennis {for two week} for two times a week/And yesterday I play basketball in {Sho} Shoin basketball club/
16P: Who do you play tennis with?
17S: /{I I play} I join to Kinki University and Shoin University's tennis circle/So circle member these days/

In the third performance, she includes most of the previous content with some variation and adds new content at 12S.

Task Performance 3
9P: Do you like to play sports?
10S: /Yeah {I like} I like watching sports and I play sports/{I} I play sports every two times a week/{I} I join

     to tennis circle for Kinki University and Shoin University {10.00} combinations circle/
11P: Where do you play?
12S: /Tennis/{I play tennis court is I I} we play tennis court in near the Kinki University's
     tennis court/{It's} it's {in} indoors tennis court/So rainy's day we can play tennis/

          The extracts demonstrate that as the student repeats the task she adds new information to the degree that
the third performance is a lot more complex than her first performance or second performance. The ability to
do this arguably comes from freed up processing capacity that is afforded by task repetition. The same can be
said for the interviewers in the study who in the latter performances appeared to be able to ask more timely
follow-up questions than in the first task performances.
          As the second and third performance extracts above show adding new information by expanding on a
previous theme or introducing a new theme came with trade-offs, as very little of the new content was error-free.
Producing this new content meant risking taking and revealed gaps in student interlanguage. Substantial or
breakdown errors were more prevalent in the third performance than the other task performances. Despite this
the third performance was still the most accurate.   The main reason for this appears that this accuracy was
formulaic-based. In other words it was not necessarily from improved syntactical processing, but more than likely
the result of students being able to employ formulaic chunks more easily without making performance errors.
Overall the success the students had in the third performance appears to be the result of eased processing
pressure that was the result of task repetition, and proximity to the second task performance, which was
immediate.

Fourth Performance
The final performance as a review was three or four weeks after the third task performance. The results for most
measures revealed that the fourth performance was generally the second best performance. An important reason
for this has to do with the relative weakness of the first performance, and the issues encountered with the second
task performance. The fourth performance more than the other three performances highlighted the difference
between the weaker and stronger students. Many of the weaker students appeared to have forgotten their prior
task work, and struggled like they had in the first performance. Many of these students spent long periods of
their fourth task performance trying to recall prior task work. In contrast, the stronger students appeared to have
developed through task repetition some familiarity with the tasks, which allowed them to recall prior language
easier as well as add new information during each task. The fourth performances appeared to show that while
task repetition temporarily facilitates improved task performances for all levels of students, the benefits appear to
deteriorate quicker for lower level students.

Implications for Classroom Practice
Transcription Suggestions
Student course-end survey feedback and lesson journal revealed that they very much enjoyed the self-correcting
part of the transcription process. One benefit of self-correcting of transcripts was that it was common for the
more proficient transcribing students to make metalanguage comments about their errors in their journals. In
contrast, most students reported that they struggled with reformulation work and did not enjoy doing it. The
findings in this study suggest that in similar contexts transcription is more suitable as 'stocktaking' at the end of a
task cycle rather than as a scaffold before a repeat task performance. It should be noted that Lynch and
Maclean's (2001) and Mennim's (2007) have reported successfully using transcription as a scaffolding tool; these
transcription studies seem to suggest it is particularly ideal for more proficient students working with language in
more restricted situational contexts. However, as the Stillwell et al. (2010) transcribing study appears to show
using transcription in these restricted contexts as well presents the same working memory issues as this study.

Task Repetition Suggestions

The most important finding for the TR component of this study was that students' production in an intact classroom improved at least temporarily in subsequent repeat task performances relative to their first performances. These results compared to the initial study also indirectly provide support that this improvement is partially dependent on having a clear pedagogical purpose supported by a strong teacher presence. For teachers using tasks, the results reveal that an initial task performance, even with pre-task planning, is probably not going to be the same quality as a repeat performance. The results from the study suggest that students with conversation tasks do rework some prior language, but overall their increased production stems from expanding on previous content or talking about something different on the same basic topic. This improved meaningfulness, or lexico-grammatical use, is perhaps the result of eased processing pressure made possible by task repetition. Finally, teachers considering using task repetition should be cautious about engaging in task repetition that resembles verbatim repetition. If meaningfulness is compromised, as was the case in the second task performances, then student task engagement seems to decrease.

Conclusion

In concluding this study, there are a number of implications for teachers who want to use TR or transcribing in similar classroom contexts. Overall, this study demonstrates that successfully actualizing pedagogical options like TR and/or transcribing in the classroom are very complex. Specifically, with transcribing, the study suggests that making the students' language more visually salient for more intensive treatment in preparation for a repeat performance does not necessarily translate into better subsequent oral task performances relative to recalling and reporting with general feedback. The task type and proficiency level appear to be important variables as well, but arguably the main variable appears to be a working memory constraint, which is probably more acute with beginner students. To date, very few transcribing studies or TR studies have acknowledged this important variable. The TR results suggest that students need to repeat the same task in order to reach the upper limits of their interlanguage. With TR it appears that a very difficult balance has to be achieved between maintaining the meaningfulness of the task and ensuring there is a transparent learning purpose supported by an equally transparent teaching presence. The final implication of this study is the complexity of the findings in terms of required classroom management and practice by the teacher highlight the need for more classroom-based studies like this one to validate the effectiveness of task-based learning in the classroom. While this study demonstrates that classroom studies can be messy, thus limited in generalizability, nonetheless they do provide teachers with a rich source of classroom-validated knowledge that can help them more fully evaluate the relevance and potential implications of key language learning concepts like task repetition and transcription for their own teaching contexts.

References

Ahmadian, M. & Tavakoli, M. (2011). Effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research,* 15, 35-60.

Bygate, M. (1996). Effects of task repetition: appraising the development language of learners. In Willis, J. & Willis, D. (Eds.), *Challenge and Change in Language Teaching* (pp, 136-146). Oxford, UK: Heinemann.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In Bygate, M., Skehan, P. & Swain, M. (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing* (pp. 23-48). Harlow: Longman.

Bygate, M. & Samuda, V. (2005). Integrative planning through the use of task-   repetition. In Ellis, R. (Ed.), *Planning and Task Performance in a Second Language* (pp.239-273). Amsterdam, Netherlands: John Benjamins.

Bygate, M. (2006). Areas of research that influence L2 speaking. In E. Uso-Juan & A. Martinez-Flor (Eds.), *Current Trends in the development and teaching of the four skills* (pp.159-186). *Berlin:* Mouton de Gruyter.

Dai, T, (2013). Working Memory as a Component of Foreign Language Aptitude. *Asian EFL Journal,* 15, 229-259.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: a unit for all reasons. *Applied Linguistics,* 21, 354–75.

Freed, B. (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder? In Riggenbach, H. (Ed.), *Perspectives on Fluency* (pp.243-265). Ann Arbor: The University of Michigan Press.

Gass, S., Mackey, A., Fernandez, M. & Alvarez-Torres, M. (1999). The effects of task repetition on linguistic output. *Language Learning,* 49, 549–80.

Hawkes, M.L. (2012). Using task repetition to direct learner attention and focus on form. *ELT Journal,* 66, 327-336.

Lynch, T. (2001). Seeing what they meant: transcribing as a route to noticing. *ELT Journal,* 55, 124-132.

Lynch T. (2007). Learning from the transcripts of an oral communication task. *ELT Journal,* 61, 311-320.

Lynch, T. & Maclean, J. (2001). Effects of immediate task repetition on learners' performance. In Bygate, M., Skehan, P., & Swain, M. (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing* (pp.141-162). Harlow: Longman.

Mennim, P. (2007). Long-term effects of noticing and oral output. *Language Teaching Research,* 11, 265-280.

Moser, J. (2008). An exploratory research study into classroom-based task repetition. Unpublished paper. University of Birmingham.

Norris, J. & Oretga, L. (2009). Towards and organic approach to investigating CAF in instructed SLA: the case of complexity. *Applied Linguistics,* 30, 555-578.

Ohta, S. (2001). *Second Language Acquisition Processes in the Classroom.* Mahwah, NJ: Lawrence Erlbaum Associates.

Pinter, A. (2005). Task repetition with 10-year-old children. In Edwards, C. & Willis, J., (Eds.), *Teachers Exploring Tasks in English Language Teaching* (pp.113-126). New York< NY: Palgrave Macmillan.

Prabhu, N.S. (1987). *Second Language Pedagogy.* Oxford: Oxford University Press.

Rossiter, M.J. 2001. The challenges of classroom-based SLA Research. *Applied Language Learning,* 12, 31-44.

Samuda, V. & Bygate, M. (2008). *Tasks in Second Language Learning.* New York, NY: Palgrave Macmillan.

Skehan, P. & Foster, P. (2005). Strategic and on-line planning: the influence of surprise information and task time on second language performance. In Ellis, R. (Ed.), *Planning and Task Performance in a Second Language* (pp.193-216).   Amsterdam, Netherlands: John Benjamins.

Stillwell. C., Curabba, B., Alexander, K., Kidd, A., Kim, E., Stone, P., & Wyle, C. (2010). Students transcribing tasks: noticing, fluency, accuracy, and complexity. *ELT Journal,* 64, 445-455.

Stones, T.P. (2013). Transcription and the IELTS speaking test: facilitating development. *ELT Journal*, 67, 20-30.

Thornbury, S. (2005). *How to Teach Speaking.* Essex, UK: Longman.

Williams, J. (2005). Form-focused instruction. In Hinkel, E. (Ed.), *Handbook of Research In Second Language Teaching and Learning* (pp.671-691). Mahwah, NJ: Lawrence Erlbaum Associates.

Willis, J. (1996). *A Framework for Task-Based Learning*. London, UK: Longman.

Willis, D. & Willis, J. (2007). *Doing Task-based Teaching.* Oxford, UK: Oxford.

Appendix

Transcript Conventions

1S: Identified learner and turn (e.g. first turn in conversation, student S)

2P: Partner and turn (e.g. second turn in conversation)

/ /   AS-unit boundaries

[ ]   Length of run boundary

{ }   Dysfluencies and pausing (not counted as part of an AS-unit)

The Teaching of Afro-Asian Literature: A Comparison between the Nonconventional Learner-Centered and the Conventional Teacher-Centered Approaches

Wilkinson Daniel O. Wong Gonzales

(wdwgonzales@yahoo.com)
De La Salle University – Manila, Philippines

Abstract

In the past few decades, nonconventional learner-centered (NLC) trends in ELT have emerged amidst the current and dominant conventional teacher-centered (CTC) approaches used by majority of private and public schools in the Philippines. The argument of which approach is more effective seems to have surfaced in the recent years. In several literatures, both approaches have been found effective in teaching in different disciplines. However, in the teaching of literature, specifically Afro-Asian literature, it seems that limited studies have been conducted in the Philippine private school context. This comparative study focuses on the Grade 8 ESL literature classes of Hope Christian High School, Manila, Philippines and aims to determine which approach is more effective in teaching Afro-Asian literature. The teacher preferred the CTC approach while students preferred the NLC approach based on qualitative data. The quantitative results showed that there were no significant differences in the Afro-Asian literature test scores immediately after the CTC or NLC instructional sessions. I conclude that there is no particular teaching approach that is more effective and suggested that an eclectic or mixed approach would be more beneficial in teaching Afro-Asian literature.

Key words: teacher-centered, learner-centered, ESL, Afro-Asian, approach, literature

Introduction

On May 15, 2013, the Enhanced Basic Education Act of 2013, more commonly known as the K-12 Program, was signed into law in the Republic of the Philippines (Official Gazette of the Republic of the Philippines, 2013). According to the Department of Education of the Philippines (DepEd) Grade 8 curriculum guide, the English curriculum for students in the eighth grade level of the K-12 Program will primarily focus on African-Asian or Afro-Asian literature (Republic of the Philippines Department of Education, 2012). By the end of the first quarter, students in the eighth grade are expected to demonstrate "understanding of the different genres through the types contributed by Afro-Asian countries to express appreciation for Afro-Asian heritage" (Republic of the Philippines Department of Education, 2012, p. 34). This ensures that the learners are exposed to Asian and African cultures through literature.

Educators teaching Afro-Asian literature to Grade 8 ESL learners must consider their reading comprehension skills. Concannon-Gibney & Murphy (2012) said that one of the goals of reading literature is comprehension, aside from the act of decoding. The ability to improve the reading comprehension of learners may determine the effectiveness of the approach utilized by the teacher. However, there are many concerns that language and literature teachers face. These concerns differ from one context to another and evolve with the diverse needs of the learners. According to Musa (2007), students in the Asian context are more likely to be deficient in comprehension and analysis skills in literature. In Japan, the standards of English reading comprehension for university students have been "steadily declining" (Nishida, 2013, p. 164). In the Philippines, one out of four people cannot read in English (Social Weather Stations, 2008). Also, a classroom analysis by Protacio & Sarroub (2013) suggested that being a good reader in a Philippine classroom is not about the reading

comprehension but rather the fluency. In other words, the importance is placed on the performance rather than the comprehension.

In the field of education, another indicator of the effectiveness of the teaching approaches an ESL teacher implements is the teacher's ability to raise the achievement test scores of his or her students (Imig & Imig, 2006). A teacher would most likely be deemed "ineffective" if he or she is unable to raise the level of student achievement in their classrooms. The Philippines used to be the top education performers in Asia; currently, it is among the lowest performers not only in Asia, but also the rest of the developing world (Raya, 2007). National Achievement Test (NAT) scores for the school year 2009-2010 for high school students dropped from 47.40% to 46.38% (Republic of the Philippines Department of Education, 2010). Statistics on an article by Raya (2007) also point out that scores on achievement tests administered by the Department of Education to Filipino students over years are "erratic and consistently low" (Raya, 2007, p. 23).

The aforementioned statistics challenge ESL teachers in the Philippines, such as those teaching Afro-Asian literature, to reassess the effectiveness of their teaching approaches so that their learners may be able to attain high scores in achievement tests, but more importantly, to actively comprehend the Afro-Asian text that they have been taught. The following section discusses the two approaches that form the focus of this study.

Conventional Teacher-Centered (CTC) Approach

In the traditional or conventional approach, the teacher is a "sage on the stage" (King, 1993). The conventional approach is called the prescriptive approach of teaching because it focuses on the teacher's monologue rather than the dialogue between the teacher and the students. In a conventional teacher-centered (CTC) classroom, the teacher is the didactic instructor and the authoritative figure (Leindhardt, 1993). The teacher transmits information instead of transforming prior knowledge. In addition, teachers who use this approach usually test the students' ability to recall information and demonstrate mastery over a narrow set of skills. The CTC approach is influenced by many theories, one of which is the Mental Discipline Theory. It advocates repetition and emphasizes learning activities that require the mind of the learners to be disciplined and trained (Tracey & Morrow, 2012). Aside from that, the teaching as telling or transmission theory of Paul Ramsden is also under the CTC approach (Ramsden, 1992). In a CTC literature classroom, ESL teachers would probably employ the lecture method. Aside from being cost-effective and economical, the lecture method is also efficient in large classrooms (Eison, 2010). According to Gibbs (2013), not all lectures involve uninterrupted presentation as student attention can be maintained. The lectures that a teacher gives in class may be interactive and are called interactive lectures (Eison, 2010; Steinert & Snell, 1999). Also, Tierney, Readence, & Dishener (1990) suggest the Listen-Read-Discuss (L-R-D) strategy where typically, fifteen minutes of direct instruction or lecture is given before the class reads the selection and the teacher guides the discussion; the emphasis of the activity is "content-driven".

The CTC approach is a very well known approach in most schools and institutions around the world since it was used before learner-centered practices became the trend (Adib-Hajbaghery & Aghajani, n.d.). According to an article, several researchers have discovered that the Philippine educational system seems to emphasize on educational goals that are related to attainment and the mastery of predefined and pre-described knowledge and skills (Bernardo, Zhang, & Callueng, 2002). The researcher, having resided in the Philippines for more than ten years and having been exposed to the CTC approach in several schools in the Philippines, attests to the fact that a lot of Philippine schools, both public and private, utilize the conventional teacher-centered model in their curriculum. Most of the time, schools in the primary to the secondary level primarily adopt this approach of teaching and expect the students to learn skills and information in a limited period of time while emphasizing mastery. Although more than 700 studies have confirmed that the lecture method is less effective, findings show that the deductive method of teaching figurative language to eighth graders produced better results on achievement test rather than the inductive, or nonconventional, method (Purves & Beach, 1972; Gibbs, 2013). According to Karagiorgi and Symeou (2005), the pre-determined conventional approach is more effective for

introductory lessons. This statement can be supported by the Bloom's Taxonomy, where the largest part of the pyramid is the Knowledge or Remembering section. In an introductory course, the teacher may choose to expose his or her students to as much knowledge they can absorb because only then can students reach the Comprehension or Understanding level. Furthermore, Michael Charleston Chua, a distinguished Philippine historian, media personality, and a notable lecturer, hypothesizes that the CTC approach may be more effective in the Philippine context. He anchored his hypothesis on the fact that Filipinos are very accustomed to the one-way oral relay of stories and past experiences, which began long before the Spanish's colonization of the archipelago (Daroy, 1969). The elders shared epics, myths, and legends to younger members of the community. Oral traditions are evident in the Philippines such in the Central Luzon province (Manuel, 1980). As a matter of fact, until the present day, these can still be observed in the Ifugao people of the Cordillera of Northern Luzon (Blench & Campos, 2010). Generally, Filipinos, as a people, are accustomed to the oral traditions. Since the one-way generation-to-generation transmission of the language is a feature of the CTC approach, the Filipinos may actually be more effectively taught in the mentioned approach since they are more accustomed to it.

Nonconventional Learner-Centered (NLC) Approach

In the nonconventional learner-centered (NLC) approach, the teacher, who provides the necessary resources, enhances the quality of discussion by allowing the students to tap into their curiosity, engage in intellectual interpersonal discourse with their peers, and encourages them to discover vital information themselves by shifting from being a "sage on the stage" to becoming a "guide on the side" (King, 1993). Instead of being the center of attention, the teacher places much importance on the enhancement of learning by giving control of learning to the students. The NLC approach is also known as the descriptive approach because it involves the aspect of interaction between students (Allwright, 1983). For this study, the approach is only considered to be nonconventional when the teacher is not the sole source of knowledge; moreover, the learners should be actively involved in the learning process. The NLC approach is influenced by theories such as Rousseau, Pestalozzi, and Froebel's Unfoldment Theory, which stresses that the students should be encouraged to allow curiosity and interests to guide one's self in learning (Tracey & Morrow, 2012). Kolb's Experiential Learning Theory is also under this approach and stresses that experience plays a major role in the learning process (Kolb, Boyatzis, & Mainemelis, 1999). Another theory that advocates the unorthodox student-centered learning is John Dewey's theory of interest, experience, and learning, more commonly known in the field of education as Inquiry Learning. Constructivism is also one of the learning theories that is categorized with this approach. Probable NLC activities implemented in a reading class would include the Sketch to Stretch strategy by Harste, Burke, Siegel, and Feathers. Educators who use this strategy requires the students to make a sketch according to their interpretations and share the sketch with the class; learning is reinforced through this activity as it does not only cover reading, but also drawing. It also takes advantage of the schema of the reader. The Jigsaw activity helps students learn cooperatively using team-learning approach; in here, a student become the expert on one part of the lesson and becomes responsible in teaching what he or she knows to other members of the group. Other NLC activities include role play and the use of buzz groups or the collaborative sharing of ideas and concepts (Tierney, Readence, & Dishner, 1990).

Despite the prevalent CTC scenario of many public and private high schools in the Philippines, Guzman (2004) highlights that the Philippines integrates the NLC approach in its "overall education framework" (Guzman, 2004, p. 223). Substantial evidence of the use of the NLC approach is found in the Philippines, typically in tertiary educational institutes such as the Polytechnic University of the Philippines, where the use of "interactive and modern strategies" is apparent (Jacolbia, n.d.). De La Salle University-Manila has a curricular framework that advocates the NLC approach. Majority of the sessions in the university comprises of learner-centered, self-exploratory, authentic, and guided activities, which is in line of the University's vision and mission statement. In a study by Weinberger and McCombs, NLC practices seem to prevail over CTC practices to improve learning; learner-centered pedagogy is a tool to improve the performance of students (Bell, 2012). In

Israel, an increase in attendance and participation in English class is evident in a student-centered reading comprehension course (Peretz, 1988). In achievement tests, students who were exposed in the NLC approach of teaching had answers that involve more classifications, comparisons, generalizations than their conventional counterparts; the Structure of the Observed Learning Outcome (SOLO) was higher, too (Tynjala, 1998). In the Philippines, the NLC approach is effective in a community college, although it does not garner high ratings based on student evaluation (Magno & Sembrano, 2007). Moreover, a significant degree of improvement in understanding can be found in students of a Philippine state university when exposed to the NLC approach compared to the CTC approach (Gravoso, Pasa, Labra, & Mori, 2008).

A Practical Framework

As illustrated in the framework below, a teacher has to prepare by having a background of the theories while being grounded on certain beliefs and equipped with learning theories to prepare lessons that may be either teacher-centered or learner-centered; some of these theories or beliefs include Unfoldment Theory, Constructivism, and Experiential Learning Theory (Tracey & Morrow, 2012; Cornelius-White, 2007). A teacher's beliefs and philosophies will influence how he or she interprets and views events to make educational decisions. In other words, the practices that are implemented by a teacher have a relationship with the personal beliefs of an educator (Gutierrez, 2004; Saroyan & Snell, 1997). According to several literatures, teachers and their teaching methods or practices seem to have an effect on reading comprehension (Samuels & Farstrup, 2011; Richardson, Anders, Tidwell, & Lloyd, 1991; Mastropieri, Scruggs, & Graetz, 2003; Yussof, Jamian, Roslan, Hamzah, & Kabilan, 2012). In addition, the type of teaching approach also has a direct affect on achievement test scores across many disciplines that include medicine, engineering, mathematics, and English (Purves & Beach, 1972; Tynjala, 1998; Terenzini, Cabrera, Colbeck, Parente, & Bjorklund, 2001; Ilyas, Rawat, Bhatti, & Malik, 2013; Khalid & Azeem, 2012). In summary, the teacher-centered or student-centered theories that an educator believes in have an effect on the practices and educational decisions that he or she makes in the classroom (the activities and strategies employed by the teacher), which affects the reading comprehension-based Afro-Asian literature achievement test scores.
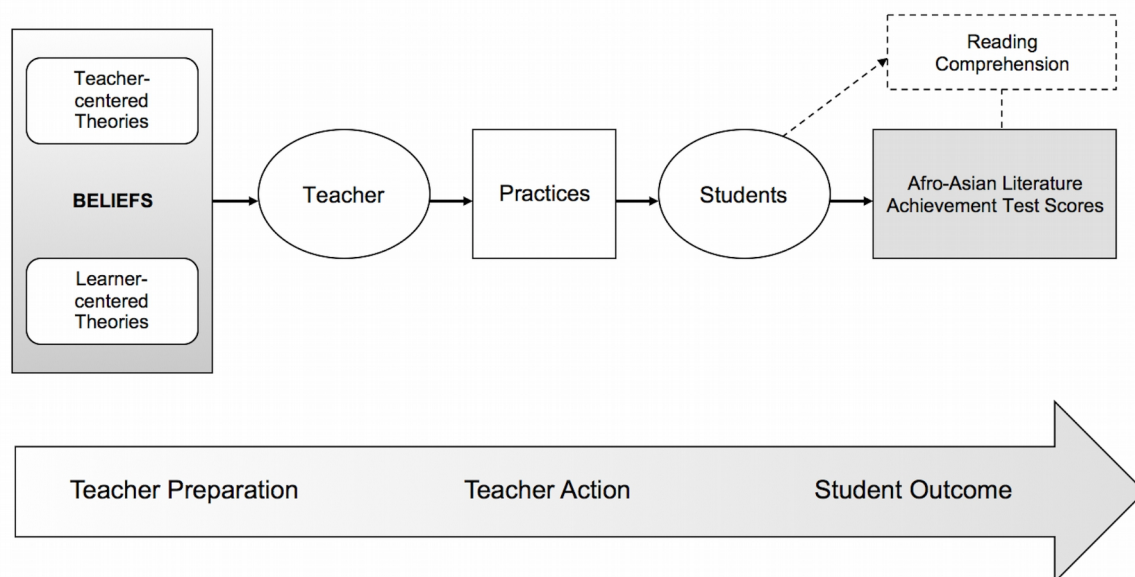

Figure 1. A Conceptual Model of the Relationship of the Choice of Theory-Influenced Approach or Approaches to Teacher Practices and Student Outcomes

Research Questions and Hypotheses

Based on the conceptual framework, literature, and results of similar studies, I aim to determine which approach would be better for teaching Afro-Asian literature. This study would make a good contribution to the ESL classroom because it seems that there are only a limited number of studies that compare the CTC and NLC approaches in the Philippine context; moreover, in the teaching of Afro-Asian literature. The study could also contribute greatly to the liberal arts subject areas.

I would like to answer the following questions: (1) Which between the CTC and NLC approaches does the teacher consider more effective in teaching Afro-Asian literature? (2) Which between the CTC and NLC approaches do the students consider more effective? And (3)Which between the CTC and NLC approaches is more effective in enhancing reading comprehension-based test scores in Afro-Asian literature?

Four null hypotheses were formulated for the purposes study and will be tested by the descriptive and inferential quantitative statistics:

H1: There is no significant difference between the CTC class and the NLC class on the standardized reading assessment/ diagnostic test.

H2: There is no significant difference between the CTC class and the NLC class on the reading comprehension-based *Asian* literature test.

H3: There is no significant difference between the CTC class and the NLC class on the reading comprehension-based *African* literature test.

H4: There is no significant difference between the CTC class and the NLC class on the reading comprehension-based *Afro-Asian* literature test.

Methodology

Research Design

I incorporated the two-group experimental mixed method of research to compare the effects of the CTC and NLC approaches of teaching on the students' reading comprehension through Afro-Asian literature achievement test scores. As stated by Creswell and Plano Clark in 2007, having a mixed method of research will provide strengths that will outweigh the weaknesses of both qualitative and quantitative research; it will also provide a better understanding of the situation (Scruggs, 2008). Similar to comparative study done by Scruggs, this method was selected to address the research questions of this study and to give a more accurate picture of the results through the use of the Convergence Model (see Figure 2).
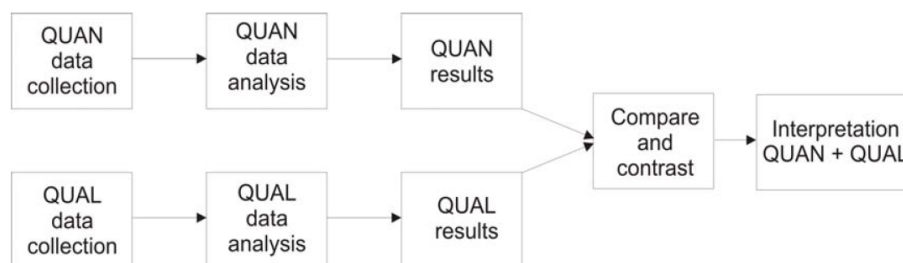

Figure 2. The Convergence Model (Creswell and Plano Clark, 2007 as cited in Scruggs, 2008, p. 36)

Quantitative data includes a) the Afro-Asian achievement test scores collected after the lessons and b) the standardized reading assessment results. Qualitative data, on the other hand, will consist of a) the responses from the resource teacher in the interview protocol and b) the observations in the observation checklist.

Participants

For the comparative study, 62 respondents belonging to the Grade 8 level with ages ranging between 13 and 15 is included in the study. CLASS A is composed of 30 students (12 female and 18 male); CLASS B is composed of 32 students (15 female and 17 male). The respondents of the research are high school students taking up Afro-Asian literature from two sections of the eighth grade level. The teacher and the respondents of this study are from Hope Christian High School, where classes are conducted in English and Filipino in the morning and Chinese Mandarin in the afternoon, which makes the institution trilingual.

Research Instruments

An observation checklist based on the review of literature is utilized and is accomplished during the observation of classes of Grade 8 students. The first part, *Teacher Behavior*, concentrates on the practices of the teacher in the classroom. The second part, *Student Behavior*, concentrates on the explicit behavior and attitude of students in the classroom. This checklist determined the emergent approach the teacher is using in the classroom and also the reaction of the students on the type of pedagogy used by the teacher. In addition, interview protocol is administered to the subject adviser. It will be used to measure the teacher's perception regarding the effectiveness of the implemented approaches: either the CTC or the NLC approach. For the study, a diagnostic test measuring reading comprehension was also administered to the respondents. The assessment contains a multiple-choice test that is adapted from Richmond, Virginia's Department of Education's Grade 8 Reading *Standards of Learning* (SOL) test released in Spring 2008. The test originally consists of 45 items focusing on analyzing printed materials, word analysis, and identifying meaning based on world literature. However, due to the constraints in time and context, slight modifications have been made. Instead of 45 items, 30 items on reading comprehension and word analysis strategies are included in the standardized reading assessment. For the purposes of this study and due to time constraint, the students are only given 30 minutes to complete the test. Furthermore, reading comprehension-based teacher-made achievement tests were given at the end of the two classroom sessions and are designed to measure the effectiveness of the approach by gauging the amount of correct items the respondents got based on his or her understanding and ability to grasp content knowledge in a typical Afro-Asian lesson. The test is based on the readings done in class and is constructed by the researcher. The 10-12 item multiple-choice test has been aligned to the learning intents of the lesson plan created by the researcher and the resource teacher.

Data Collection

Figure 3 outlines the data collection procedure. First, I conducted an interview with the teacher and asked her the questions in the interview protocol. After the interview, the teacher provided the weekly lesson plans. After that, I made and suggested four lesson plans—2 NLC lesson plans and 2 CTC lesson plans—to the cooperating teacher. I used the lesson plans to construct two multiple choice reading comprehension-based achievement tests on the topics: The *Tale of Genji* by Lady Murasaki and *Things Fall Apart* by Chinua Achebe. Also, I prepared the booklets and answer sheets on the standardized reading assessment, which will be used as a diagnostic test.
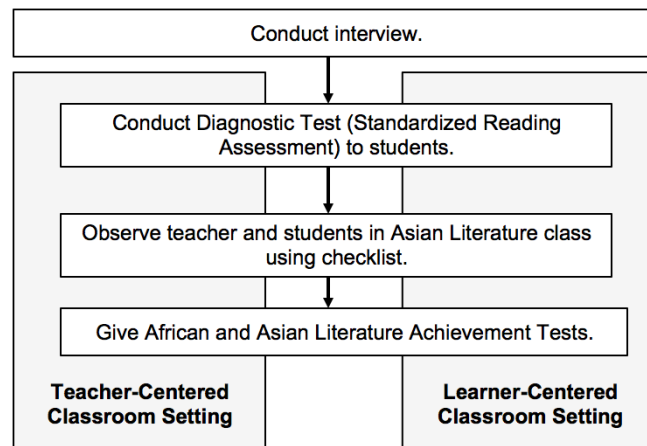
Figure 3. The Process of Administering the Research Instruments

During the students' homeroom period, the diagnostic test was administered to the students to measure their reading comprehension and analysis skills. After the diagnostic test, observation of eighth grade Afro-Asian literature classes on both sections was conducted. The observation took place in two classrooms. The same teacher taught Japanese Literature (Asian Literature representative) to the two sections the same material using the CTC approach and the NLC approach respectively. The class that is taught using the CTC method, for the purposes of this study, will be referred to as CLASS A, and the other group, CLASS B. CLASS A was exposed to the L-R-D procedure where two-way communication is minimized. CLASS B, on the other hand, was taught using the NLC approach and was exposed to an inductive discussion, jigsaw group activity, and synthesis; interaction was encouraged. During the classes, I utilized an observation checklist to record classroom practices of teacher participants. After the lessons, I gave a 10-item multiple choice reading comprehension-based achievement test that is related to the Asian literature lesson. Fifty minutes were allotted for each class.

After two weeks, I asked the cooperating teacher to execute another set of lesson plans focusing on African literature. The CTC and NLC approach is again used on CLASS A and CLASS B, respectively. CLASS A was taught using the L-R-D procedure while CLASS B was taught using collaborative activities such as buzz groups. CLASS B was also instructed to make a group collage on their understanding of the novel. After that, a 12-item multiple-choice test was administered. Fifty minutes were allotted for each class.

Analytical Procedure
Descriptive and inferential analysis was conducted on all quantitative data. SPSS Version 22 for Macintosh was used to analyze the data along with Microsoft Excel 2011 for Macintosh.

Results

Qualitative Data
Interview Protocol
The resource teacher strongly believes that a combination of approaches would be for the best. Since she advocates the injection of culture in literature class, what she usually does is give a summary (direct instruction) and then give learner-centered activities (indirect instruction). However, she also believes that currently, the CTC should be more emphasized compared to the NLC approach. She thinks it would be more effective for the Grade 8 students because they do not have much prior knowledge on the literary works she presented. She believes that the effectiveness of an approach is determined by how well the objectives of the day are met.

Observation Checklist

CLASS A. In the class that was taught using the CTC approach, the teacher gave a lecture, was the center of attention, and acted as the "sage on the sage". The teacher attempted to discipline those who did not behave by telling them to keep quiet and asked that they raise their hands if they wanted to talk. In addition, it was observed that the teacher transmits the information instead of transforming prior knowledge. During the class period, the teacher lets the students work alone. However, the teacher was lenient and allowed the free sharing of ideas. The lecture was a little interactive. The teacher also contextualized the activity or discussion based on the interests of the students.

Less than half of the students were engaged in the discussion. Only some were enthused by the topic of the day. Most of the students were misbehaved; however, the content of their discourse is highly intellectual. There are instances in class where I caught two students sleeping in class. I also observed that none of the students were taking notes during the class.

CLASS B. In the class that was taught using learner-centered strategies, the teacher gave a group work and encouraged group members to collaborate. I observed that the teacher is the "guide on the side" and serves as the facilitator. Although she could have just shared the story to the students, the teacher allowed the exploration of the material and the free expression of ideas while contextualizing the activities. It can be seen from the observation period that the teacher tries to make the students the source of information by making them experts of the content they hold during the Jigsaw activity. The teacher also tries to discipline the class by making them keep silent and telling them to raise their hands whenever they want to ask a question.

All of the students were enthusiastic about the topic, but not all are engaged. I observed that due to the lack of resources and space, some of the students were not able engage in the Jigsaw activity. Most of the time, the students are not behaved and are caught talking to each other. However, I listened to them closely and observed that their discourse is intellectual. As a matter of fact, the quality of questions that were asked was in fact very critical. No one in the class slept. None of the students in the class that are taught using the NLC approach were taking notes.

Quantitative Data

Standardized Reading Assessment / Diagnostic Test

Administered Test	Sig.
Diagnostic Test	0.946
Asian Literature Achievement Test	0.012
African literature Achievement Test	0.622
Afro-Asian literature Achievement Test (averaged mean scores)	0.013

*Figure 4.*The Significance Values of the Scores Based on Levene's Test for Equality of Variances

The descriptive and inferential statistics showed that the mean score of CLASS A (10.967) and CLASS B (11.469) in the standardized reading assessment has a significance value of 0.946 based on Levene's Test for Equality of Variances (see Figures 4 and 5). Since it is above 0.05, I interpreted the results assuming that the variances were approximately equal. The variance for CLASS A has a value of 35.62 while CLASS B has 30.13. The t-test indicates that p-value 0.731 is insignificant at $\alpha=0.05$ level of significance (see Figure 7). The first null hypothesis states that there is no significant difference between the CTC class and the NLC class on the standardized reading assessment or diagnostic test. The hypothesis is, therefore, accepted, since the t-value is greater than 0.05 (Figure 7). Since there is no significant difference between the CTC class and the NLC class on the standardized reading assessment or diagnostic test, it is safe to assume that both CLASS A and CLASS B were equal based on their diagnostic test scores in reading comprehension.
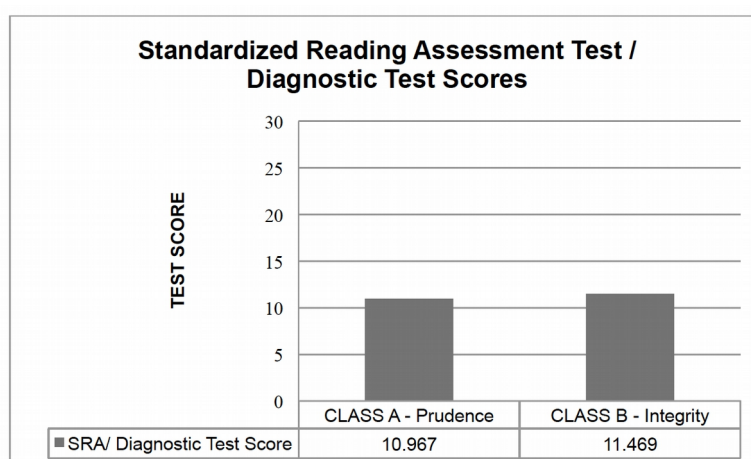

Figure 5. Standardized Reading Assessment Test Scores for Selected Grade 8 Students
in Hope Christian High School

Reading Comprehension-based Achievement Tests
The second null hypothesis of the study was that there is no significant difference between the CTC and NLC classes in Asian literature test. The results showed a difference between the mean scores of the CTC class (6.167) and the NLC class (4.938) (see Figure 6). The significance value in Levene's Test for Asian literature test scores was 0.012 (see Figure 4). Since it is below 0.05, I interpreted the results assuming that the variances were not equal. The significance of the difference was tested using t-test for independent samples t-statistics revealed that p-value 0.051 is insignificant at $\alpha=0.05$ level of significance (see Figure 7). Thus, the null hypothesis stating that there is no significant difference between CLASS A and CLASS B on their Asian literature test is accepted, although it was evident that the performance of the CTC class is slightly better than the performance of the NLC class.

The third null hypothesis of the study was that there is no significant difference between the CTC class and the NLC class in the African literature achievement test. The results showed a difference between the mean scores of CLASS A (4.167) and CLASS B (4.188) (see Figure 6). The variance for CLASS A has a value of 2.971 while CLASS B has a value of 4.222. The significance value in Levene's Test was 0.622 (see Figure 4). Since it is above 0.05, I interpreted the results assuming that the variances were approximately equal based on Levene's Test. T-statistics revealed that p-value 0.966 is insignificant at $\alpha=0.05$ level of significance. So the null hypothesis that states that there is no significant difference between the CTC and NLC classes is accepted. Therefore, there is no statistically significant difference between the CTC and NLC class in the African literature test.
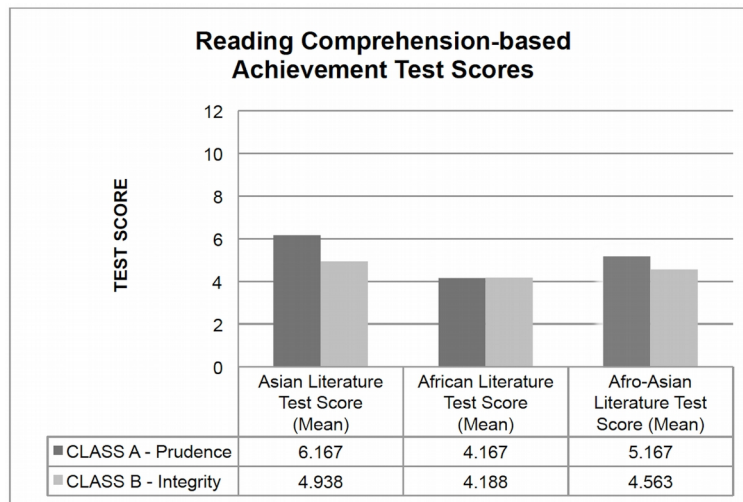
Figure 6. Reading Comprehension-based Test Scores of Selected Grade 8 Students
in Hope Christian High School

The fourth null hypothesis of the study was that there is no significant difference between the CTC and NLC classes in Afro-Asian literature. Since the study focuses on Afro-Asian literature and not independently Asian or African literature, I averaged the test scores of both CLASS A and CLASS B on their Asian and African tests. The averaged mean score of the CTC class is 5.167 unlike the averaged mean score of CLASS B, which is 4.563 (see Figure 6). The significance value in Levene's Test for Afro-Asian literature test scores was 0.013 (see Figure 4). Since it is below 0.05, I interpreted the results assuming that the variances were not equal. The significance of the difference was tested using t-test for independent samples t-statistics revealed that p-value 0.135 is insignificant at α=0.05 level of significance (see Figure 7). Thus, the null hypothesis stating that there is no significant difference between CLASS A and CLASS B's Afro-Asian literature test score is accepted, although it was evident that the performance of the CTC class in Afro-Asian literature is slightly better than the performance of the NLC class (see Figure 6).

Since the reading comprehension of both classes in the diagnostic test is equal according to Levene's Test for Equality of Variances, it is safe to assume that the t-test for the averaged means of the Afro-Asian literature test scores between the CTC and NLC classes is valid.

Hypothesis	Cl.	n	Mean	σ^2	SD	Mean Diff.	t	Sig. (p)	Decision
Ho1: There is no significant difference between the CTC class and the NLC class on the standardized reading assessment/diagnostic test.	A	30	10.967	35.62	5.97	0.502	.345	0.731*	ACCEPTED
	B	32	11.469	30.13	5.49				
Ho2: There is no significant difference between the CTC class and the NLC class on the reading comprehension-based *Asian* literature test.	A	30	6.167	8.005	2.83	1.229	-2.000	0.051*	ACCEPTED
	B	32	4.938	3.544	1.88				
Ho3: There is no significant difference between the CTC class and the NLC class on the reading comprehension-based *African* literature test.	A	30	4.167	2.971	1.72	0.021	0.043	0.966*	ACCEPTED
	B	32	4.188	4.222	2.05				
Ho4: There is no significant difference between the CTC class and the NLC class on the reading comprehension-based *Afro-Asian* literature test.	A	30	5.167	3.023	1.72	0.604	-1.517	0.135*	ACCEPTED
	B	32	4.563	1.851	1.36				

*Significance level α = 0.05

Figure 7. Hypotheses Testing.

Discussion

There are several limitations to this study. One possible limitation is that some of the students were absent during the diagnostic test and the achievement tests. A score of 0 was given to those who cannot attend. This could be a limitation, as the scores may be extremely affected. Another possible limitation concerns the teacher-made tests in this study, as it may not accurately assess the students' reading comprehension in Afro-Asian literature, given that I am not an expert in creating reading comprehension assessments. In addition, time was very limited because I was only able to conduct observations on specific days, as allotted by the partner teacher. Furthermore, the number of respondents, both teachers and students, may also be a limitation of the study.

Despite these shortcomings, this study provides some interesting findings. From the qualitative data, more specifically the interview protocol, the teacher advocates the combination of the CTC and NLC approaches. However, she believes that the CTC approach should be more dominant compared to the NLC approach because from her experience, students with no prior knowledge are most likely clueless in her learner-centered activities. Her belief is consistent with the Benjamin Bloom's Taxonomy model. In relation to the review of related literature, how can students reach comprehension or understanding without knowing? The resource teacher also mentioned that for an approach to be considered effective, it has to meet the objectives of the lesson.

If the objective of the teacher is to recognize characters in the story without any prior knowledge, then the teacher should opt to use the CTC approach. If the objective of the teacher is to let the students infer the plot of the story considering they have prior knowledge already, then the NLC approach would be deemed better. In connection to the findings of the interview protocol, if, according to the teacher, the learning objectives of the teacher are not accomplished under the utilized NLC approach, it should only mean that the approach is not that effective. However, this does not imply that the NLC is not effective while the CTC approach should be used for all settings. CTC can be used in selected scenarios. In the study by Karagiorgi and Symeou (2005), the CTC approach is more effective in introductory lessons.

Aside from the fulfillment of the teacher's objectives of the lesson, another factor that should be considered to determine the effectiveness of an approach is the interest of the students during the implementation of the approach. When a teacher implements a lesson using either the CTC or NLC approach and it sustains the students' attention for the duration of the lesson, then the utilized approach is considered effective by the students. From the qualitative data, more specifically the results of the observation checklist, it was discovered that in the CTC class, some of the students were not enthusiastic about the lesson at all. There were two students who were seen sleeping during the class. On the other hand, in the NLC class, all of the students were enthusiastic about the lesson of the day. No one was seen sleeping during the lesson. This shows that the students prefer the NLC approach over the CTC approach.

From the quantitative data, which includes the results of the standardized reading assessment/diagnostic test and the achievement test scores in Afro-Asian literature, several conclusions can be made. In Asian literature, although the data suggests that there is no significant difference, there is a minimal difference between the scores. The CTC class had a slightly higher mean score compared to the NLC class. This suggests that the conventional approach may appear to be slightly effective in teaching Asian literature. However, based on the data analyses, the choice of approach seems to have no bearing on the reading comprehension-based Asian literature test scores of the students. In African literature, there is also no significant difference between the test scores of the CTC and NLC classes. Again, this just proves that the type of approach used by the teacher may not have an effect on the reading comprehension-based African achievement test scores of the students. However, since the focus of this paper is in the Afro-Asian context, and independently assessed results may not effectively answer the question at hand, it was imperative to average the mean scores of the African and Asian literature tests of both classes. I discovered that there is no significant difference in the reading comprehension-based Afro-Asian literature test scores between the classes even after considering averaging the mean scores. Therefore, from the data analyses, it can be concluded that the choice of the CTC or the NLC approach has no significant effect in raising reading comprehension-based Afro-Asian literature test scores in Hope Christian High School. The findings of this study were consistent with other researches on the type of approach used (Chung, 2004; Chilwant, 2012; Ottman, n.d.; Hafezimoghadam, Farahmand, Farsi, Zare, & Abbasi, 2013).

Conclusion

For the study, triangulation was used so that the conclusions would be more comprehensive and reliable. The effectiveness of the CTC and NLC approach was looked at from the perspective of the student, teacher, statistics, and the review of related literature. According to teacher, the combination of approaches would be most effective as long as the objectives are met and as long as one knows when to use the approach. In the context of Hope Christian High School, the CTC approach would be more effective for the teacher since the students may lack prior knowledge. For the students, the NLC approach is more effective because it sustains their interest towards the lesson. From the statistical data, the choice of the CTC or the NLC approach has no significant effect in raising reading comprehension-based Afro-Asian literature test scores. Therefore, from the data analyses and the perspectives of students and teachers, I conclude that there is indeed no particular approach that is more effective in teaching Afro-Asian literature in Hope Christian High School, Manila, Philippines.

ESL teachers may use an eclectic approach instead of the pure CTC or NLC approach in teaching Afro-Asian literature. They may use the CTC approach for introducing lessons and vocabulary. They may also opt to use it to discuss the lesson; after all, the CTC approach may be interactive. Apart from using the CTC approach, educators may also use the NLC approach in Grade 8 Afro-Asian literature classes. After the lesson, the teacher may utilize NLC reading comprehension strategies such as the buzz groups, Jigsaw, or role play to improve the performance of the students. Using an eclectic approach may be more effective compared to an independent approach. Perhaps, the degree of learner-centeredness or teacher-centeredness is one factor that the teacher should consider in her classroom when he or she teaches Afro-Asian literature. Nevertheless, a mixed approach may be better than merely using either the CTC or NLC approach, especially in the teaching of Afro-Asian literature.

References

Adib-Hajbaghery, M., & Aghajani, M. (n.d.). *Traditional lectures, socratic method and student lectures: Which one do the students prefer?* Retrieved February 7, 2014, from Web Med Central: http://www.webmedcentral.com

Allwright, D. (1983, June). Classroom-centered research on language teaching and learning: A brief historical overview. *TESOL Quarterly , 17*(2), 191-204.

Bell, C. D. (2012). *The Relationship of Learner-Centered Beliefs of Principals and Student Achievement*. Ann Arbor, Michigan: ProQuest LLC.

Bernardo, A. B., Zhang, L. F., & Callueng, C. M. (2002). Thinking styles and academic achievement among Filipino students. *The Journal of Genetic Psychology*, 149-163.

Blench, R., & Campos, F. (2010). Recording oral literature in a literate society: a case study from the Northern Philippines.

Chilwant, K. S. (2012). Comparison of two teaching methods, structured interactive lectures and conventional lectures. *Biomedical Research*, 363-366.

Chung, I. (2004). A comparative assessment of constructivist and traditionalist approaches to establishing mathematical connections in learning multiplication. *Education, 125*(2).

Concannon-Gibney, T., & Murphy, B. (2012). Reading comprehension instruction in Irish primary classrooms: key insights into teachers' perspectives on classroom practices. *Irish Educational Studies*, 433-449.

Cornelius-White, J. (2007, March). learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research, 77*(1), 113-143.

Daroy, P. (1969). The novel in the Philippines. *Asian Studies, 7*(2), 180-185.

Eison, J. (2010). using active learning instructional strategies to create excitement and enhance learning. 1-20.

Gibbs, G. (2013, Novemeber 21). *Lectures don't work, but we keep using them*. Retrieved February 7, 2014, from Times Higher Education: http://www.timeshighereducation.co.uk/news/lectures-dont-work-but-we-keep-using-them/2009141.article

Gravoso, R., Pasa, A., Labra, J., & Mori, T. (2008). design and use of instructional materials for student-centered learning: A case in learning ecological concepts. *The Asia-Pacific Education Researcher*, 109-120.

Gutierrez, M. C. (2004). Studying teacher cognition: The interplay of teacher beliefs and instructional practice. *Kritika Kultura*, 53-78.

Guzman, A. (2004). The hermeneuticcs of learner-centered approaches and initiatives in the Philippine basic education sector. *Educational Research for Policy and Practice*, 223-241.

Hafezimoghadam, P., Farahmand, S., Farsi, D., Zare, M., & Abbasi, S. (2013). A comparative study of lecture and discussion methods in the education of basic life support and advanced cardiovascular life support for medical students. *Klinik Calisma*, 59-63.

Ilyas, B., Rawat, K., Bhatti, M., & Malik, N. (2013, January). effect of teaching algebra through social constructivist approach on 7th grader's learning outcomes in Sindh (Pakistan). *International Journal of Instruction, 6*(1).

Imig, D. G., & Imig, S. R. (2006). The teacher effectiveness movement: How 80 years of essentialist control have shaped the teacher eduction profession. *Journal of Teacher Education*, 167-180.

Jacolbia, R. B. (n.d.). Preferred and effective learner-centered strategies at the Polytechnic University of the Philippines. *International Conference: The Future of Education*.

Khalid, A., & Azeem, M. (2012). Constructivist vs traditional: Effective instructional approach in teacher education. *International Journal of Humanities and Social Science, 2*(5), 170-177.

King, A. (1993). From Sage on the Stage to Guide on the Side. *College Teaching, 41*(1), 30-35.

Kolb, D. A., Boyatzis, R. E., & Mainemelis, C. (1999, August 31). The evolution of John Mezirow's transformative learning theory.

Leindhardt, G. (1993). On teaching. (R. Glaser, Ed.) *Advances in Instructional Technology*, 1-54.

Magno, C., & Sembrano, J. (2007). The role of teacher efficacy and characteristics on teaching effectiveness, performance, and use of learner-centered practices. *The Asia Pacific Education Researcher*, 73-91.

Manuel, E. A. (1980). Philippine oral traditions: Theory and practice. *Philippine Quarterly of Culture & Society*, 7-27.

Mastropieri, M. A., Scruggs, T. E., & Graetz, J. E. (2003). Reading comprehension instruction for secondary students: Challenges for struggling students and teachers. *Learning Disability Quarterly , 26*(2), 103-116.

Musa, M. A. (2007). A comparative study between English literature and Asian literature through the perceptions of the TESL Foundation Programme students. *Seminar Penyelidikan Pendidikan Institut Perguruan Batu Lintang Tahun*, 1-17.

Nishida, H. (2013). The influence of chunking on reading comprehension: Investigating the acquisition of chunking skill. *The Journal of Asia TEFL*, 163-183.

Official Gazette of the Republic of the Philippines. (2013, May 15). *Republic Act No. 10533*. Retrieved 6 2013, November, from Official Gazette of the Republic of the Philippines: http://www.gov.ph/2013/05/15/republic-act-no-10533/

Ottman, L. (n.d.). The effect of student-directed versus traditional teacher-centered presentations of content on studetn learning in a high school statistics class. Haddon Heights, New Jersey, United States of America.

Peretz, A. (1988). Student-centered learning through content-based instruction: use of oral report projects in the Advanced EFL Reading Class. *Reading in a Foreign Language*, 181-192.

Protacio, M., & Sarroub, L. (2013). A case study of reading instruction in a Philippine classroom. *Asia Pacific Journal of Education*, 338-349.

Purves, A. C., & Beach, R. (1972). *Literature and the Reader: Research in response to literature, reading interests, and the teaching of literature.* Urbana, Illinois: National Council of Teachers of English.

Ramsden, P. (1992). *Learning to teach in higher education.* New York: Routledge.

Raya, R. R. (2007). 'the missed education of the Filipino people' missing targets: an alternative MDG midterm report. *Social Watch Philippines*.

Republic of the Philippines Department of Education. (2010, October). Discussion paper on the enhanced K+12 Basic Education Program: DepEd discussion paper. Philippines.

Republic of the Philippines Department of Education. (2012, December). K to 12 curriculum guide: English (Grades 7-8). Pasig, Philippines.

Richardson, V., Anders, P., Tidwell, D., & Lloyd, C. (1991). The relationship between teachers' beliefs and practices in reading comprehension instruction. *American Educational Research Journal, 28*(3), 559-586.

Samuels, S., & Farstrup, A. E. (2011). *what research has to say about reading instruction.* Newark, Delaware: International Reading Association.

Saroyan, A., & Snell, L. (1997). Variations in lecturing styles. *Higher Education*, 85-104.

Scruggs, B. B. (2008). *Learning outcomes in two divergent middle school string orchestra classroom environments: A comparison of a learner-centered and a teacher-centered approach*. Atlanta, Georgia: ProQuest LLC.

Social Weather Stations. (2008, May 16). *First quarter 2008 social weather survey: National proficiency in English recovers*. Retrieved August 26, 2014, from Social Weather Stations: http://www.sws.org.ph/pr080516.htm

Steinert, Y., & Snell, L. (1999). Interactive lecturing: strategies for increasing participation in large group presentations. *Medical Teacher*, 37-43.

Terenzini, P., Cabrera, A., Colbeck, C., Parente, J., & Bjorklund, S. (2001, January). Collaborative learning vs. lecture/discussion: Students' reported learning gains. *Journal of Engineering Education*, 123-130.

Tierney, R. J., Readence, J. E., & Dishner, E. K. (1990). *Reading strategies and practices: A compendium* (3rd Edition ed.). Massachusetts: Allyn and Bacon.

Tracey, D. H., & Morrow, L. M. (2012). *Lenses on reading: An Introduction to theories and models.* New York: The Guilford Press.

Tynjala, P. (1998). Traditional studying for examination versus constructivist learning tasks: Do learning outcomes differ? *Studies in Higher Education*, 173.

Yussof, Y., Jamian, A., Roslan, S., Hamzah, Z., & Kabilan, M. (2012). Enhancing reading comprehension through cognitive and graphic strategies: A constructivism approach. *Procedia Social and Behavioral Sciences*, 151-160.

English Language Anxiety among Arab Postgraduate Users: A Case of Arab Postgraduates in a Higher Education Institution in Malaysia

Asmaa AlSaqqaf
(asma3030@yahoo.com)
Siti Jamilah Bidin
Ahmad Affendi Shabdin
School of Education and Modern Languages, Universiti Utara Malaysia (UUM), Malaysia

Abstract

This study aimed to examine English anxiety among Arab postgraduates studying at a higher education institution in Malaysia within an ESL (English as a Second Language) context. It also investigated whether the Arab postgraduates' demographic variables could affect their level of language anxiety. Horwitz, Horwitz and Cope's (1986) concepts of communication apprehension and fear of negative evaluation together with Gardner's (2004) English use anxiety constituted the theoretical framework of the present study. Horwitz et al.'s (1986) Foreign Language Classroom Anxiety Scale (FLCAS) was adapted and used in this study together with the construct of English Use Anxiety in Gardner's (2004) Attitude/ Motivation Test Battery. Based on two-dimensional constructs reflecting communication within the academic contexts and outside in everyday communication situations, this study quantitatively surveyed and analyzed 122 Arab postgraduates studying at a higher education institution in Malaysia. Results indicated that Arab postgraduates experienced a slightly high anxiety while using English in the academic contexts, with a moderate level of anxiety when speaking English in everyday communication situations. Furthermore, communication apprehension and fear of negative evaluation were reported to be the two salient types of anxiety which Arab postgraduates experienced the most. Findings also revealed that gender and length of stay in Malaysia were found to have significant differences in Arab postgraduates' language anxiety, whereas age and level of study (MA and PhD) hardly had any impact in this regard. Finally, the possible causes leading to such findings were also discussed.

Key words: English language anxiety, Arab postgraduates, academic context, gender.

Introduction

Anxiety is considered as one of the most significant affective factors that can seriously influence foreign language (FL)/ second language (SL) acquisition. Spielberger (1983) defines anxiety as the "subjective feeling of tension, apprehension, nervousness, and worry associated with an arousal of the automatic nervous system" (as cited in Horwitz et al., 1986, p.125). Due to their fear of making mistakes or losing face while speaking in English, FL/SL anxious students usually speak with shaking hands, highly beating heart, and some can even go blank when having to speak in an FL/SL. Thus, many choose to avoid using English by remaining silent in class.

Much research has been conducted to examine the scope and severity of language anxiety among FL/ SL learners. However, the vast majority of this research deals with classroom based anxiety (Woodrow, 2006), and hence, the focus was only on school and undergraduate students. Therefore, it is not surprising to discover that hardly any investigation has been carried out to examine language anxiety use outside the classroom among language users, and not only learners. Remarkably, very limited research has been conducted on Arab students and postgraduates in this regard. Additionally, the studies that have investigated the relationship between language anxiety and some demographic variables such as gender, age, and length of time spent in the target

language country, have yielded mixed results. Therefore, this article attempts to fill this gap by examining English language anxiety among Arab postgraduates studying at a higher education institution in Malaysia based on two-dimensional constructs reflecting communication within and outside academic contexts in everyday communication situations.

Foreign Language Anxiety

Anxiety is classified into three types: state, trait and situational. State anxiety takes place within specific and temporary situations and fades when these situations (or threats) come to an end (MacIntyre & Gardner, 1991a; & Spielberger &Vagg, 1995). Trait anxiety, on the other hand, refers to the individual's trait of becoming anxious in any situation (Scovel, 1978; & Spielberger, 1983). Situational anxiety, however, is a specific form of anxiety that occurs consistently over time within a given situation (MacIntyre & Gardner, 1991a). Since foreign language anxiety is prompted by a specific set of conditions such as public speaking or participating in class (Ellis, 2008), Horwitz et al. (1986) and Gardner (1985a) distinguish FLA from the first two categorizations of anxiety, and classify it as a situation-specific anxiety.

Horwitz, Horwitz and Cope's (1986) study has been influential in theorizing and measuring anxiety in relation to a foreign language. Horwitz et al., (1986) theorize foreign language anxiety as comprising three components: communication apprehension, fear of negative evaluation, and test anxiety.

Communication apprehension is a type of shyness associated with fear of or anxiety about communicating with others. It includes "difficulty in speaking in dyads or groups (oral communication anxiety) or in public (stage fright), or in listening to or learning a spoken message (receiver anxiety)" (p.127). Horwitz et al. (1986) argue that communication apprehension has a dominant role among the three other constructs of foreign language anxiety.

Test anxiety, on the other hand, is related to the tension or emotional interruption caused by fear of failure. Students with test anxiety tend to set high-performance goals that can be difficult to achieve, and thus put much pressure on themselves. For sensitive and vulnerable students, simultaneous occurrence of tests and oral communication is highly likely to make the phenomenon of test anxiety become obvious (Horwitz et al., 1986).

Fear of negative evaluation refers to the uneasiness and tension about other people's evaluation, especially criticism. It is broader in scope than the concept of test anxiety, which is limited to test situations as it can occur in any evaluative circumstances such as interviewing for employment or speaking a foreign language in class. Fear of negative evaluation is always associated with those sensitive students who are doubtful about their abilities in language classes.

From Horwitz et al.'s (1986) study emerged the 33 item FLCAS that has been widely used in a great number of studies (Horwitz, 2001). The scale has been found to be reliable and valid (Aida 1994; & Cheng, Horwitz & Schallert 1999). Since then, studies on foreign language anxiety using the FLCAS have flourished.

Research on language anxiety among postgraduates obviously suffers from an extreme dearth. Among the very few studies in this area is Brown's (2008) study that investigated international postgraduates' anxiety over their level of the English language at a university in the south of England. Brown reported that although all students enrolled in their course with a minimum level of IELTS Band 6, the majority felt disadvantaged by particularly poor spoken English, and suffered feelings of anxiety, shame and inferiority. Low self-confidence, as found by Brown (2008), meant that students felt ill-equipped to engage in class discussion and in social interaction which used English as the medium of communication. Brown also observed that a common reaction to stress caused by language problems was to retreat into monoethnic communication with students from the same country, further inhibiting progress in language. In addition, Brown reported that whilst some 'linguistic progress' was made by nearly all students during the academic sojourn, the anxiety suffered by students in the initial stage must not be underestimated, and appropriate support systems must be put in place to alleviate their distress.

Generally, students with communication apprehension, test anxiety, and fear of negative evaluation are

highly likely to develop language anxiety with different levels which, in turn, can be attributed to some demographic variables such as age, gender, length of time spent in the target language country, and perceived language ability. Researchers have explored the relationship between foreign language anxiety and those background variables, but the studies yielded mixed results.

For example, gender is one of the background variables often discussed in relation to foreign language anxiety. However, this relationship is still not clearly established in the literature due to the mixed results. For instance, Alidoost, Mirchenari and Mehr (2013) examined the effect of anxiety on 102 high schools learners of English in Iran. Results showed that there was no significant difference between male and female learners in most anxiety variables. Also, the findings revealed that students experienced some degree of anxiety in English classes, and indicated that the main sources of anxiety for Iranian students were second language deficits, fear of negative evaluation, and perfectionism.

Elkhafaifi (2005) examined the language anxiety among 233 graduate and undergraduate students of Arabic as a foreign Language enrolled in Arabic language programs at 10 universities in the United States. Results demonstrated that advanced students had lower language anxiety compared to beginners or intermediate students. The length of language learning was also identified as a second factor in language anxiety as findings revealed that older students (sophomores, juniors, and seniors) who had spent more years learning English in school experienced lower anxiety in comparison with younger students (freshmen). In addition, Elkhafaifi (2005) found that female students showed higher anxiety compared to male students. In contrast, Matsuda and Gobel (2004) who investigated language anxiety among 252 university students majoring in English found no significant effect of gender on students' anxiety.

Huang (2005) explored speaking anxiety among 502 EFL university students in Taiwan. The results differed significantly among the students according to gender and time of starting to learn English. While females were shown to experience a higher speaking anxiety about EFL learning compared to males, different time of starting to learn English was also reported to have a significant difference in language anxiety. On the other hand, Huang (2005) found that age hardly had any significant difference in language anxiety among students of different ages, in contrast to Ohata (2005) who claimed that age played a role in language anxiety.

Li's (2010) study revealed that 309 non-English freshmen and sophomores in Taiwan suffered from language anxiety. However, no statistically significant differences in speaking anxiety between male and female students were detected. However, starting to learn English was found to have a significant difference in language anxiety.

Casado and Dereshiwsky (2001) examined the levels of anxiety as experienced by first and second semester university students speaking Spanish as a foreign language. They observed that students' level of language anxiety increased slightly rather than decreased with more exposure to language learning.

As seen above, studies investigating the impact of gender, age, and the length of exposure to the target language on language anxiety have yielded mixed results. While Alidoost et al (2013), Li (2010), and Matsuda and Gobel (2004) concluded that gender had no significant difference in language anxiety, other studies affirmed the impact of gender but with contradicting results. For example, Abu-Rabia (2004), Elkhafaifi (2005), and Huang (2005) reported that female students showed higher anxiety compared to male students, whereas Na (2007) and Shi and Liu (2006) revealed that male students were more anxious than their female counterparts .Similarly, length of exposure to the target language was also found to be associated with foreign language anxiety, but the results are also still inconclusive. While some studies (Elkhafaifi, 2005; Huang, 2005; Li, 2010; & Casado & Dereshiwsky, 2001) demonstrated that the length of exposure to the target language greatly reduced students' anxiety level, other studies (e.g. Caruso, 1996) reported that there was no significant difference in this regard.

Hence, further research of how language anxiety changes with these demographic variables is needed. The purpose for investigating the relationship between these background variables and foreign language anxiety is not only to determine about the relationship itself, but also to deepen the understanding of the nature of

foreign language anxiety in terms of knowing why foreign language anxiety level is related to or not related to these variables. Furthermore, much of language anxiety research has been conducted on school students or undergraduate learners of FL/SL in classroom contexts. Limited research has been carried out on FL/SL users outside classroom contexts and with students of different backgrounds such as Arab postgraduates. Therefore, reducing these gaps justified the present study to be conducted.

Research Questions

In an attempt to reduce the gap identified in the literature, the present study investigates English language anxiety among Arab postgraduates based on two-dimensional constructs reflecting communication within and outside the academic contexts in everyday communication situations. To achieve this purpose, this study attempts to answer the following questions:

1. Do Arab postgraduates studying in Malaysia experience anxiety while using English in academic contexts?
2. Do Arab postgraduates studying in Malaysia experience anxiety while using English outside the academic context in everyday communication situations?
3. What types of anxiety do Arab postgraduates experience the most?
4. Is there any relationship between the demographic variables (i.e. gender, age, level of academic study, and length of stay in Malaysia) of Arab postgraduates and their level of anxiety?

Methodology

Participants

The study involved 122 Arab postgraduates, 26 of whom were female students, whereas 96 were males. In terms of their academic study, 62 were PhD students while 60 were Masters students. A demographic profile of the respondents is presented in Table 1.

The Instrument

As mentioned earlier, Horwitz et al.'s (1986) FLCAS consists of 33 items using a five-point Likert Scale, which is widely used in empirical studies of language anxiety. The measure, focusing on speaking in an FL learning context, investigates learners' communication apprehension, test-anxiety and fear of negative evaluation. Given that the present study's respondents who are studying in a higher education institution in Malaysia (where English is the medium of instruction) are considered English users rather than English learners, some items of the FLCAS were modified to suit the characteristics of the respondents. For example, the item I never feel quite sure of myself when I am speaking in my foreign language class was modified to become I feel unsure of myself when I am speaking English in my classes/ presentations/ seminars/ workshops / etc.

Yet, there were still some items which did not suit the respondents of this study even with the help of a considerable modification. That is because those items represent the context of learning English as a subject rather than the context of using English as a medium of instruction (e.g., the item I worry about the consequences of failing my foreign language class). Accordingly, such items had to be excluded. As a result, 19 items were found to be suitable for the respondents of the present study. In addition, the five items under the construct English Use Anxiety in the Gardner's Attitude/ Motivation Test Battery (the 2004 version) were also added to the survey because they deal with the use of English outside the classroom, which matched the characteristics of respondents. Thus, these 24 items constituted the instrument of the present study, and were labeled as English Use Anxiety Scale (EUAS). The EUAS was initially translated by the first co-author of the present paper (being a native speaker of Arabic language as well as a specialist in English), and then sent to two external reviewers who were also native speakers of Arabic as well as specialists in English, for the purpose of evaluating the naturalness and accuracy of the translated version.

Table 1
Distribution of the Respondents by Demographic Variables (n=122)

Gender	Frequency	%
Females	26	21.3
Males	96	78.7
Study Level		
Masters	60	49.2
PhD	62	50.8
Age Group		
Less than 25 to 30	51	41.8
More than 30 to 35	49	40.2
More than 35 to 40	18	14.8
More than 40 to 45	4	3.3
Duration of Stay in Malysia		
1 month – 1 year	51	41.8
More than 1 year – 2 years	41	33.6
More than 2 years – 3 years	15	12.3
More than 3 years – 4 years	15	12.3
Nationalities	Frequency	%
Libyan	20	16.4
Iraqi	50	41.0
Jordanian	30	24.6
Syrian	2	1.6
Egyptian	2	1.6
Saudis	2	1.6
Alegerian	9	7.4
Palestinian	5	4.1
Lebanese	2	1.6
Major		
Applied Linguitics	11	9.0
Information Technology	51	41.8
Business Management	34	27.86
Accounting	13	10.65
Banking & Finance	8	6.55
Law	5	4.09

After being translated, the Arabic version of the survey was distributed to the participants who were requested to rate each item on a five-point Likert Scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). 350 survey questionnaires were distributed to PhD and Masters students in a higher education institution in

Malaysia. 131 copies were retrieved, out of which 9 were found invalid due to the incompletion of some sections of the scale. Therefore, only 122 copies were subject to analysis.

Data Analysis

SPSS (Statistical Product and Service Solutions) 18.0 was used to analyze the data obtained. Before proceeding to the analysis, the data had been prepared for the analysis through employing a number of basic procedures, i.e., cleaning and screening the data, handling the missing values, outlier detection, and a normality testing through checking the skewness and kurtosis of the data obtained.

Noticeably, some studies (e.g., Burden, 2004; Cheng, 2009; Liu & Huang, 2011; & Na, 2007) categorize Horwitz et al.'s (1986) FLCAS items into three constructs for the purpose of representing the three areas (i.e., communication apprehension, fear of negative evaluation and test anxiety) highlighted in Horwitz et al.'s (1986). However, it can be argued that this categorization seems to lack a statistically clear base. Consequently, the first co-author of the present study corresponded with E. K. Horwitz via an e-mail to inquire about this trend followed by some studies. Horwitz mentioned that:

> In my opinion the FLCAS does not have 3 dimensions, so there is no way to differentiate it into 3 dimensions. Horwitz, Horwitz, and Cope 1986 only state that CA, test anxiety, and fear of negative evaluation are related to foreign language anxiety. In fact they argue that foreign language anxiety is different from those three constructs. If a researcher believes that there are different dimensions to the FLCAS, that hypothesis could be tested with factor analysis.
>
> (E. K. Horwitz, personal communication, July 15, 2013)

Following Horwitz's suggestion, an exploratory factor analysis (EFA) was performed to the 19 items adopted from Horwitz et al.'s (1986) FLCAS in order to identify the latent structures (dimensions) of the asset variable through the use of the Principle Component Analysis (PCA) with Varimax rotation.

Results of Factor Analysis

EFA basically aims to define the underlying structures among variables in the analysis (Hair et al., 2010). However, the suitability of the data for factor analysis is firstly examined through two tests, i.e., Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy (MSA), and Bartlett's test of sphericity. The recommended value of KMO is .60 or above (Kaiser, 1970; Kaiser, 1974), whereas the significance value of Bartlett's test of sphericity should be .50 or smaller (Pallant, 2001). Table 2 shows the results of the factor analysis conducted for the 19 items.

Table 2 shows that KMO is .932, highly exceeding the recommended value of 0.60, and that Bartlett's test of sphericity is statistically significant at p = 0.000. These results indicated that the data was appropriate for factor analysis. The factor loadings of the items were examined and compared with the minimum value of 0.50 recommended by Hair et al. (2010) for practically significant item loading. As can be seen from the data in Table 2 above, the values of the factor loading ranged between 545 and .842, exceeding the recommended limit of 0.50 for practical significance. The 19 items loaded on two factors with Eigen values greater than one. These two factors explained 63.14 % of their overall variance, and showed a high internal reliability of 0.946 and 0.890 for the first and second factors respectively. Based on the common meaning and content of the items grouped under each factor, the factors were labeled (Hair et al., 2010). Communication Apprehension (CA) was named to the first factor, whereas the second factor was labeled as Fear of Negative Evaluation (FNE).

Table 2
Factor Analysis for the 19 items adopted from Horwitz et al. (1986)

Item Code	Factor Loading	
	CA*	FNE**
AN10	.793	
AN9	.788	
AN13	.769	
AN11	.769	
AN12	.717	
AN14	.716	
AN2	.697	
AN1	.678	
AN6	.666	
AN8	.643	
AN5	.623	
AN4	.564	
AN7	.545	
AN16		.842
AN18		.794
AN17		.701
AN19		.674
AN15		.592
AN3		.558
Eigen value	10.819	1.178
VE %	56.941	6.201
Reliability	0.946	0.890
KMO	0.932	
Chi square	1756.770	
Significance	0.000	
* CA	Communication Apprehension	
** FNE	Fear of Negative Evaluation	

Methods of Analysis

To answer the questions of the present study, the mean and standard deviation (SD) were calculated to determine whether Arab postgraduates feel anxious while speaking English in the academic contexts as well as in everyday communication situations, and also to identify what scopes of anxiety Arab that postgraduates experience anxiety the most. Furthermore, the mean, standard deviation as well as the independent-samples t-test were run to

identify any significant difference in language anxiety among male and female students, as well as Masters and PhD students. In addition, a one-way between-groups analysis of variance (ANOVA) was conducted to explore the impact of age groups and the length of stay in Malaysia on levels of foreign language anxiety.

Results

Q.1 Do Arab postgraduates studying in Malaysia experience anxiety while using English in academic contexts?

The results of the descriptive analyses showed that 15 out of the 19 items used to measure the respondents' language anxiety in the academic contexts were found to have values above 3.00. The means of the items ranged from 2.56 (item 16) to 3.42 (item 6) with SDs ranging from 1.034 (item 5) to 1.256 (item 17). Item 16 had the lowest mean (M = 2.56) which indicated that almost 55 % (67 out of 122) of the participants disagreed or strongly disagreed that the more they prepare for their presentations, the more confused they get. Item 6, on the other hand, had the highest mean (M = 3.42) which showed that 53.3% (65 out of 122) of the participants agreed or strongly agreed that they start to panic when they have to speak in English without preparation during their classes/ seminars/ workshops/ etc. Table 3 below shows a summary of these results.

Table 3
Arab postgraduates' average English anxiety experienced in the academic contexts (N=122) (based on the results obtained from a five-point Likert Scale)

Measure	No. of items used	No. of items with values above 3	Lowest mean	Highest mean	Average language anxiety
Items adapted from Horwitz et al.'s FLCAS ,1986)	19 items	15 items	2.56 (item 16) SD=1.106	3.42 (item 6) SD=1.163	3.10 (rounded)

In conclusion, the responses revealed that the extent to which average language anxiety among Arab postgraduates was slightly higher (M = 3.99, 58.89 /19) than median value of 3, which indicated that the respondents experienced a slightly high language anxiety while using English in the academic contexts.

Q.2 Do Arab postgraduates in Malaysia experience anxiety while using English in everyday communicative situations?

The results obtained from the construct of English use anxiety were different from those of the CA and FNE constructs that measured language anxiety in the academic contexts. Results shown in Table 4 below demonstrated that the means of the items ranged from 2.26 (item 24) to 2.64 (item 22) with SDs ranging from 1.011 (item 24) to 1.143 (item 22). Obviously, the responses obtained demonstrated that Arab postgraduates experienced a moderate anxiety (M = 2.48, 12.14/5) lower than the median value 3.

Table 4
Arab postgraduates' average English anxiety experienced in the everyday communication contexts (N=122) (based on the results obtained from a five-point Likert scale)

Measure	No. of items used	No. of items with values above 3	Lowest mean	Highest mean	Average language anxiety
English Use Anxiety (adopted from Gardner's AMTB, 2004 version)	5 items	Nil	2.26 (item 24) SD=1.011	2.64 (item 22) SD=1.143	2.48

Q.3 What types of anxiety do Arab postgraduates experience anxiety the most?

The computation of means and standard deviations of the three constructs of anxiety demonstrated that the respondents' communication apprehension reached a mean of 3.18, which is considered high. This indicated that the Arab postgraduates experienced communication apprehension the most, followed by fear of negative evaluation with a mean of 2.92. English use anxiety had the lowest mean among the three constructs with a mean of 2.48. Table 5 below shows the types of anxiety that Arab postgraduates were reported to experience.

Table 5
Descriptive data of the three constructs of language anxiety as experienced Arab postgraduates

Anxiety Types	Mean	SD
Communication Apprehension (CA)	3.18	1.127
Fear of Negative Evaluation (FNG)	2.92	1.191
English Use Anxiety (EUA)	2.48	1.067

Q. 4 Is there any relationship between the demographic variables (i.e. gender, age, level of academic study, and length of stay in Malaysia) of Arab postgraduates and their level of anxiety?

Relationship between Gender and Level of Anxiety

The means and standard deviations of the 24 items constituting the instrument of the present study indicated that the Arab female postgraduates were found to be highly anxious with a mean reaching 3.41 compared to the mean of 2.85 for their male counterparts as shown in Table 6 below. Interestingly, 22 items (from item 1 to item 22) out of the 24 ones were rated by the females with scores above 3.00. Generally, the means of all the items rated by the females ranged from 2.58 (item 24) to 3.85 (items 1 and 2) with SDs ranging from .736 (item 12) to 1.350 (item 17).

As far as the Arab male postgraduates are concerned, only 11 items out of 24 had scores above 3, where item 5 had the highest mean of 3.33, while item 24 had the lowest mean of 2.14, with SDs ranging from1.000 (item 21) 1.206 (item 17). In addition, the results of the independent-samples t-test presented in Table 7 below reported a significant difference in the scores between the males (M= 81.85, SD=18.46) and the females (M= 68.44, SD=19.51) with a t-value (df 120) =3.142, and the result was significant at p=0.002< .05.

Table 6
Descriptive data of anxiety levels between the Arab females and males

Gender	Mean	SD
Female (n=26)	3.41	1.091
Male (n=96)	2.85	1.113
Total (n=122)	2.97	1.130

Table 7
Results of the independent-samples t-test for gender

Variable	Gender	Levene's Test for Equality of Variances		T-test for Equality of Means		
EUAS	Equal variances assumed	F	Sig.	t	df	Sig. (2-tailed)
		.437	.510	3.142	120	.002

Relationship between Age Group and Level of Anxiety
A one-way between-groups ANOVA was conducted to determine the impact of age on levels of language anxiety. Respondents were divided into four groups according to their age as shown previously in Table 1. As shown in Table 8 below, the F value indicated that there was no significant difference at the $p < .05$ level ($F = 1.372$, $p=. 255> .05$) among the different age groups. In other words, the Arab postgraduates' level of anxiety did not differ significantly based on their age groups.

Table 8
Results of ANOVA for age groups

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1630.946	3	543.649	1.372	.255
Within Groups	46764.431	118	396.309		
Total	48395.377	121			

Relationship between Level of Study (Master and PhD) and Level of Anxiety
With regards to the Masters and PhD students, the means and SDs of the three constructs showed that there was no significant difference in the level of anxiety between the Masters and PhD students. The t-test also supported this finding as the t value [df 118.441 (due to the unequal variances)] was -.205 with a significance value at $p=0.838>.05$. Tables 9 and 10 present these results.

Table 9
Descriptive data of FLAUS between Masters and PhD students

Level of Study	Mean	Standard Deviation
Master (n=60)	2.96	1.120
PhD (n=62)	2.99	1.141
Total (n=122)	2.97	1.130

Table 10
Results of the independent-samples t-test for level of study

Variable	Gender	Levene's Test for Equality of Variances		T-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
EUAS	Equal variances assumed	4.430	.037	-.205	120	.838
	Equal variances notassumed			-.205	118.441	.838

Relationship between Length of Stay in Malaysia and Levels of Anxiety
A one-way between-groups ANOVA was also conducted to investigate the impact of length of stay in Malaysia on levels of language anxiety. Respondents were divided into four groups according to their length of stay in Malaysia (Group 1: 1 month - 1 year; Group 2: more than 1 year - 2 years; Group 3: more than 2 years - 3 years; and Group 4: more than 3 years - 4 years). As shown in Table 11 below, the results of ANOVA indicated that there was a statistically significant difference at the $p< .05$ level for the four groups [$F (3, 118) = 5.155$, $p=.002$]. The effect size, calculated using eta squared, was .116, which is considered as a medium effect based on Cohen's (1988) classification of .01 as a small effect, .06 as a medium effect, and .14 as a large effect-as great.

Table 11
Results of ANOVA for length of stay in Malaysia

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5607.332	3	1869.111	5.155	.002
Within Groups	42788.045	118	362.611		
Total	48395.377	121			

Post- hoc comparison using the Tukey HSD test indicated that the mean scores for both Group 1 (M=74.29, SD= 19.04) and Group 2 (M=76.12, SD=19.07) were significantly different from Group 4 (M=57.86, SD=16.25). Group 3 (M=61.33, SD=21.35) did not differ significantly from the other three Groups 1, 2, and 4.

In conclusion, the findings obtained from investigating the impact of Arab postgraduates' demographic variables (i.e., gender, age, level of academic study, and length of stay in Malaysia) on their level of language anxiety reported that gender and length of stay in Malaysia were found to have significant differences on Arab postgraduates' language anxiety, whereas age and level of stay (MA and PhD) hardly had any impact in this regard. Table 12 shows a summary of the findings for Research Question 4.

Table 12
Summary of the Finding for Research Question 4

Respondents' Background Variables	Method of Analysis	Significant Difference	(p<0.05)
Gender	*t*-test	Yes	0.002*
Age	ANOVA	No	0.255
Level of Study (MA & PhD)	*t*- test	No	0.838
Length of Stay in Malaysia	ANOVA	Yes	0.002*

Discussion
Arab Postgraduates' Experience Anxiety while Using English in Academic Contexts
Similar to the studies conducted by Brown (2008) on postgraduate students, and by Li (2010), Burden (2004), and Huang (2005) on university students, as well as those carried out by Alidoost et al. (2013) and Na (2007) on high schools students, this study revealed that the Arab postgraduates studying at a higher education institution in Malaysia experienced feelings of language anxiety, particularly in the academic contexts as communication apprehension was reported to be the salient type of anxiety suffered the most by the postgraduates. Such results can be interpreted based on many aspects. Firstly, this anxiety can be attributed to the postgraduates' self-perceived lack of English proficiency. Arab postgraduates believe that their English proficiency is not adequately good as required by the tertiary-level contexts, and feel ill-equipped to engage in class discussion. These feelings and beliefs cause them to feel anxious to communicate freely or express their viewpoints in the academic contexts which require a good linguistic competence on the part of the students that can enable them to criticize, analyze and argue upon debatable or controversial topics.

Secondly, Arab postgraduates understand that the tertiary-level contexts sometimes require them to communicate in English for a relatively longer time in order for them to express their viewpoints or argue upon topics. This likelihood of talking in English for an extended period of time is believed by Arab postgraduates to open the door for possible grammatical mistakes, wrong pronunciation and the like to occur. Thus, those perceived possibilities cause those postgraduates to develop a sense of communication apprehension while speaking English in the academic contexts, and thus they become more anxious compared to speaking outside the academic contexts.

Thirdly, the feeling of anxiety is also likely to be attributed to the perceived image of a postgraduate student in the minds of Arabs. Actually, there is a common belief among Arabs that postgraduate students, who are believed that they have reached a higher level of education, must have a high level of English proficiency. In other words, for Arabs, a postgraduate's value lies in his/ her English proficiency. If his/her English is high, then a high value is given to him/her. Actually, such a high belief does not encourage students. Rather, it often results in causing more anxiety, indeed. That is because such a belief also implies that students with low English proficiency do not deserve to be a postgraduate; hence their fear of communication because communication is a tool through which an evaluation of language proficiency can be made. Consequently, Arabs postgraduates have become convinced that better to remain silent and be thought as a good student than to speak and to remove all doubt. As a result, Arab postgraduates become very concerned about their performance in front of their peers, and thus get more anxious when they have to speak in formal contexts because they try not to 'corrupt' their image nor have any feeling of shame in front of others.

Importantly, communication apprehension among Arab postgraduates can also be explained within the context of their fear of negative evaluation. As mentioned above, Arab postgraduates believe that the value of a postgraduate lies in his/ her English proficiency. The communication apprehension is developed as a result of their fear of negative evaluation because communication is a medium through which a language evaluation can be made. This means that the students would be perceived as a good student based on the quality of their communication and language proficiency. Therefore, they experience communication apprehension because they do not want to be labeled as bad students in the minds of their peers.

Fourthly, the presence of many other international students, particularly those with relatively high English proficiency (e.g. Nigerians), makes Arab students draw a lot of comparison between themselves and those international students, concluding that other international students are better at English than them. Thus, this perception of having a lower linguistic competence compared to other international students can cause Arab students to suffer from feelings of inferiority and low self-confidence, leading them to display language anxiety in the academic contexts.

A fifth reason can be attributed to the general emphasis on all postgraduates to produce grammatical-error-free pieces of postgraduate works and assignments. Although this is considered as a good practice, it makes students pay full attention to submitting a good assignment, while less emphasis is given to improve their communication skills.

However, the situation outside the academic contexts is different. The students experience a moderate anxiety because they feel relaxed as a result of their awareness that they are not being evaluated by lecturers or peers. In addition, they feel that local people, even if they are good in English, do not have the professional English language characterizing the postgraduate academic contexts, hence the feeling of relaxation. Moreover, the location of the respondents' higher education institution in the northern part of Malaysia, where English can be considered as a foreign language and few people can speak English, enhances the feeling of low anxiety among Arab postgraduates, and thus they freely communicate in English.

Arab Female Postgraduates are more Anxious than Males

As mentioned previously, research on language anxiety shows inconclusive results regarding the role of gender in language anxiety. The results of the present study are in consistent with those obtained by Abu-Rabia (2004), Elkhafaifi (2005), and Huang (2005) reporting that the female students show higher anxiety compared to the male students. In addition to the explanations mentioned above regarding the sources of communication apprehension and fear of negative evaluation experienced by Arab postgraduates in general, the Arab female postgraduates' higher anxiety can be attributed to other additional reasons, the most important of which is the role of culture.

It is a well-known fact that Arab culture is still characterized by a male dominance, where a prevailing attitude and a culture of discrimination against women still exist in the Arab region (Al Maaitah et al., 2011).

Women have a minor role in life and are perceived to be in the second row as "the personal sphere focuses on women responsibilities for the family in the household, while the public sphere is the men's world" (Al Maaitah et al., 2011). This accounts for the dominant presence of men in almost all aspects of life in those societies. This has generated a feeling in Arab women that priority, even to speak, should be given to men, particularly in formal settings. This cultural practice has been negatively reflected in their life, especially in the educational settings, leading to a high anxiety of speaking in the presence of men. As a result of these cultural beliefs, it may not be an exaggeration if it is claimed that women not only have a high anxiety, but also a low level of confidence especially in the presence of men. This accounts for the females' fear to communicate in English in the postgraduate contexts.

Another reason is also related to the psychological nature of Arab females. Sense of shyness is one of the characteristics that distinguishes Arab females especially in the activities that require participation in the public. This feeling of shyness may cause the females to feel reluctant to be involved in communicative interaction, and consequently this feeling can be developed to communication apprehension.

A third reason can be given to the small number of female postgraduates studying in Malaysia compared to the males, which makes them feel that they are a minority whose participation may not be appreciated especially with the highly dominant presence of males. All these reasons work together and lead to a high level of anxiety among the Arab females compared to the males.

No Impact of Age and Level Study on Anxiety Compared To the Length of Exposure to English

The age of 100 respondents (i.e., 82% of the respondents) range between 25-30 and 30-35 years old. Therefore, it can be argued that there is no serious difference in the range of age between these two groups. In other words, the two groups can be considered as one generation. On the other hand, Master or PhD level of study refers to an individual's level of knowledge in his/ her area of specialization rather than his/her knowledge or proficiency in English. For further clarification, there can be a Masters student who, upon having been exposed to English for two years or more, can have an English proficiency better than a PhD student who has just come to Malaysia for doing his/her PhD and thus considered as a freshman as to English use. Therefore, in both cases (age and level of study) the concern is about the level of anxiety, regardless the age or level of study. Therefore, both these demographic variables should be explained within the context of the length of exposure to English.

As affirmed by MacIntyre and Gardner (1991a), "as experience and proficiency increase, anxiety declines in a fairly consistent manner" (p. 111). In other words, the postgraduate students who have been living for three to four years in Malaysia must have been using English during this period, and consequently gained a good linguistic knowledge in terms of learning new vocabularies, grammar rules, as well as developing their speaking and communication skills whether within or outside the academic contexts in everyday communication situations. This positive effect on the students' linguistic competence undoubtedly develops a sense of confidence while using English anywhere, something which results in low anxiety. However, the newcomers (whether Masters or PhD students) who have spent between one month to one year or more still lack confidence, and thus feel communication apprehension due to the reasons mentioned above. Accordingly, it can be affirmed that it is not a matter of age or study level. Rather, it is a matter of being exposed to English, the variable that can develop a sense of confidence that, in turn, leads to low anxiety.

Conclusion

The present study reveals that language anxiety can be considered as a phenomenon that is not associated only to a particular group of individuals such as school or university students. Thought of as highly educated people who are adult and matured enough to overcome psychological issues, Arab postgraduates were reported to feel anxious while speaking English, particularly in the academic contexts. Communication apprehension and fear of negative evaluation were the two salient kinds of anxiety which Arab postgraduates experienced the most. Gender and duration of being exposed to English were found to have an influence on the level of anxiety among

Arab students. In the light of the results of this study, further empirical research on the impact of language anxiety on other speaking-related variables (e.g., willingness to communicate, communication strategies) needs to be conducted for the purpose of getting a deeper understanding of this phenomenon as well as exploring its influence on language acquisition.

Finally, this study suggests that Malaysian higher education institutions are advised to take into consideration the results of this study in order to help international postgraduates, whether Arabs or those of other nationalities and backgrounds, to improve not only their English proficiency, but also to include some modules in the training programs of the higher education institution concerned, especially for the freshmen or newly registered students to increase their confidence in their language abilities.

References

Abu-Rabia, S. (2004).Teachers' role, learners' gender differences, and FL anxiety among seventh-grade students studying English as a FL. *Educational Psychology, 24*(5), 711-721.

Aida, Y. (1994). Examination of Horwitz, Horwitz, and Cope's construct of foreign language anxiety: The case of students of Japanese. *Modern Language Journal, 78*(2), 155-168.

Alidoost, Y., Mirchenari, N., &Mehr, S. (2013). The effects of anxiety on Iranian EFL learners' performance on final achievement test in English classes of high schools. *International Journal of Language Learning and Applied Linguistics World4*(3), 284-292.

Al Maaitah, R., Al Maaitah, H., Olaimat, H., & Gharaeibeh, M. (2011). Arab women and political development. *Journal of International Women's Studies*, *12*(3), 7-26.

Brown, L. (208). Language and anxiety: An ethnographic study of international postgraduate students. *Evaluation & Research in Education 21*(2), 75-95

Burden, P. (2004). The teacher as facilitator: Reducing anxiety in the EFL university classroom. *JALT Hokkaido Journal*, *8*, 3-18.

Casado, M.A., & Dereshiwsky, M.I. (2001). Foreign Language Anxiety of University Students. *College Students Journal, 35*(4), 539-551.

Caruso, C. L. (1996).*Foreign language anxiety and selected learner variables in adolescent language learners*. Unpublished doctoral dissertation, University of Wisconsin, Madison.

Cheng, C. (2009). *Language anxiety and English speaking proficiency.* Unpublished master thesis, Ming Chuan University, Taiwan.

Cheng, Y.S., Horwitz, E.K., & Schallert, D.L. (1999). Language anxiety: Differentiating writing and speaking components. *Language Learning, 49*(3), 417-446.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.).Hillsdale, New Jersey: Erlbaum.

Elkhafaifi, H. (2005). Listening comprehension and anxiety in the Arabic language classroom. *The Modern Language Journal, 89*(2), 206-220.

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford, UK: Oxford University Press.

Gardner, R. C. (1985a). *Social psychology and second language learning: The role of attitudes and motivation*. London, UK: Edward Arnold.

Gardner, R. C. (1985b). *The Attitude/Motivation Test Battery: Technical Report*. Ontario, Canada: University of Western Ontario.

Hair, J.F., Anderson, R.E.,Tatham, R.L.& Black, W.C.(2010).*MultivariateDataAnalysis*.(7th ed.). Upper Seddle River, NJ: Prentice Hall.

Horwitz, E. K. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics*, *21*, 112-126.

Horwitz, E. K., Horwitz, M. B., &Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal*, *70*, 125-32.

Huang, H. W. (2005). *The relationship between learning motivation and speaking anxiety among EFL non-English major freshmen in Taiwan.* Unpublished master thesis, University of Technology, Chaoyang, Taiwan.

Kaiser, H., F. (1970). A second generation little jiffy. *Psychometrika*, *35*, 401-415.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31-36.

Li, R. (2010). *The relationship between speaking anxiety and speaking strategies among university students in Taiwan*. Unpublished master thesis, National PingTung University of Education, Taiwan.

Liu, M., & Huang, W. (2011). An exploration of foreign language anxiety and English learning motivation. *Education Research International.* Retrieved October, 2013 from http://www.hindawi.com/journals/edri/2011/493167/abs/

MacIntyre, P. D. & R. C. Gardner. (1991a). Methods and results in the study of anxiety and language learning: A review of the literature. *Language Learning, 41*(1), 85-117.

Matsuda, S., & Gobel, P. (2004). Anxiety and predictors of performance in the foreign language classroom. *System, 32*, 21-36.

Na, Z. (2007). A study of high school students' English learning anxiety. *Asian EFL Journal, 9*(3), 22-34

Ohata, K. (2005). Potential sources of anxiety for Japanese learners of English: Preliminary case interviews with five Japanese college students in the U.S. *TESL-EJ, 9*(3), 1-21.

Pallant, J. (2001). *SPSS survival manual: Step by step guide to data analysis using SPSS*. Berkshire: Open University Press.

Woodrow, L. (2006). Anxiety and speaking English as a second language. *RELC Journal, 37*(3), 308-328.

Scovel, T. (1978). The effect of affect on foreign n language learning: A review of the anxiety research. *Language Learning, 28*, 128-142.

Shi, Y. Z., & Liu, Z. Q. (2006). Foreign language reading anxiety and its relationship to English achievement and gender. *Journal of PLA University of Foreign Languages, 29*, 59-65.

Spielberger, C. D. (1983). *Manual for the state-trait anxiety inventory.* Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D, &Vagg, P. (Eds.). (1995). *Test anxiety: A transactional process model, test anxiety theory, assessment, and treatment*. Washington, D.C.: Taylor & Francis.

Appendix A

English anxiety scale adapted from Horwitz et al.'s (1986) FLCAS

1. I feel unsure of myself when I am speaking English in my classes/presentations/ seminars/ workshops / etc.

2. I worry about making language mistakes while speaking English in my classes/ presentations/ seminars/ workshops / etc.

3. I am afraid that the other students will laugh at me when I speak English.

4. I keep thinking that the other students are better at English than I am.

5. I get nervous when I do not understand every word my lecturer/ supervisor /examiner says.

6. I start to panic when I have to speak English without preparation during my classes/ seminars/ workshops/ etc.

7. I always feel that the other students speak English better than I do.

8. I feel self-conscious about speaking English in front of other students

9. I get nervous when I am speaking English in my classes/presentations etc.

10. I feel overwhelmed by the number of rules I have to use to speak English.

11. It frightens me when I do not understand the English that my lecturer/supervisor is saying in classes/ meetings etc

12. I tremble when I know that I am going to be called on in my classes/ seminars/workshops etc.

13. I can get so nervous in my class/presentation etc., that I forget things I know.

14. It embarrasses me to volunteer answers in my class/ seminars/ workshops etc.

15. Even if I am well prepared for my presentations etc, I feel anxious about it.

16. The more I prepare for my presentations, the more confused I get.

17. I can feel my heart pounding when I am going to be called on in my classes/conferences/workshop.

18. I often feel like not going to my classes/conferences etc in which I have to give a presentation

19. I get nervous when my lecturers ask questions which I have not prepared in advance.

Appendix B

English Use Anxiety adopted from Gardner's (2004) Attitude/Motivation Test Battery

1. I would get nervous if I had to speak English to a tourist.

2. Speaking English anywhere makes me feel worried.

3. It would bother me if I had to speak English on the telephone.

4. I would feel uncomfortable speaking English anywhere.

5. I feel anxious if someone asks me something in English.

2015 TESOL International Conference

Organized by *TESOL Asia*, *TESOL International Journal* and
School of Foreign Languages, Shanghai University, Shanghai, China

The 2015 TESOL International Conference will be held at Shanghai University (Baoshan Campus) from May 14 to 16, with the theme of "TESOL in the Global Age: New Theories and Methodologies." Six leading linguists and scholars in the fields of linguistics and English education, Rod Ellis, Stephen Krashen, Michael Byram, Wen Qiufang, Gao Yihong, and Chen Jianlin, will be invited to make keynote speeches at the Conference. Also invited to the conference will be scholars from language education presses and academic institutions. This Conference aims to provide a high-level platform for teachers of institutions of higher learning to exchange their research findings and explore opportunities and ways to meet challenges in the globalization of education.

Main topics of the conference
1. Frontier theories in foreign language teaching
2. Language testing and assessment
3. Modern education technology and English teaching
4. Syllabus and course design
5. China's English language teaching: Reform and way-out

Conference dates May 14-16, 2015
Registration on May 14 at # 1 New Lehu Building, Shanghai University (Baoshan Campus)

Working language English

Registration fee
1200 RMB (900 RMB for students); 200 US dollars for international participants (150 US dollars for students)

Conference venue
The TESOL international conference will be held at Shanghai University (Baoshan Campus), 99 Shangda Road, Baoshan District, Shanghai, China. Please visit http://shf.shu.edu.cn for more information about the School.

Important contact information
Dr. Zhu Yiner: 15021783500 Xu Yang: 18818219916
Office telephone: +86-21-66133047
Fax: +86-21-66133058 E-mail: tesolshu2015@hotmail.com
Address: 99 Shangda Road, Baoshan District, School of Foreign Languages, Shanghai University, Shanghai, China   Zip code: 200444

Further information of the conference will be published on the conference website soon and details will be circulated.

2015 TESOL 国际研讨会通知

　　由 *TESOL Asia*、*TESOL International Journal* 和上海大学外国语学院主办的 2015 TESOL 国际研讨会将于 2015 年 5 月 14 日-16 日在上海大学宝山校区举行，主题是"全球化背景下的 TESOL：新理论与新方法"。本次研讨会将邀请国内外六位英语教育领域的知名学者 Rod Ellis、Stephen Krashen、Michael Byram、文秋芳、高一虹和陈坚林作主题报告。大会还邀请语言教育出版机构和学术机构的学者参会。本次国际研讨会旨在搭建一个平台，供广大高校英语教师交流学术思想，共同探讨全球化对英语教育教学带来的机遇与挑战。

一、研讨会的议题：
（1）外语教学理论前沿
（2）语言测试与评估
（3）现代教育技术与英语教学
（4）教学大纲与课程设置
（5）中国英语教育教学改革与出路
二、会议时间及报到地点：2015 年 5 月 14 日-16 日（14 日报到，上海大学宝山校区乐乎新楼 1 号楼）
三、会议使用语言：英语
四、会议费用：会务费 1200 元（学生 900 元）；外籍参会者 200 美元（学生 150 美元）
五、会务组联系方式：
　　联系人：朱音尔 15021783500　　徐旸 18818219916
　　办公室电话：021-66133047　传真：021-66133058
　　大会筹备组邮箱：tesolshu2015@hotmail.com
　　通讯地址：上海市宝山区上大路 99 号 上海大学外国语学院
　　邮编：200444
　　网址：http://shf.shu.edu.cn

详细信息会见网站更新，敬请关注。热忱欢迎国内外学者参加本次 TESOL 国际研讨会！
上海大学外国语学院