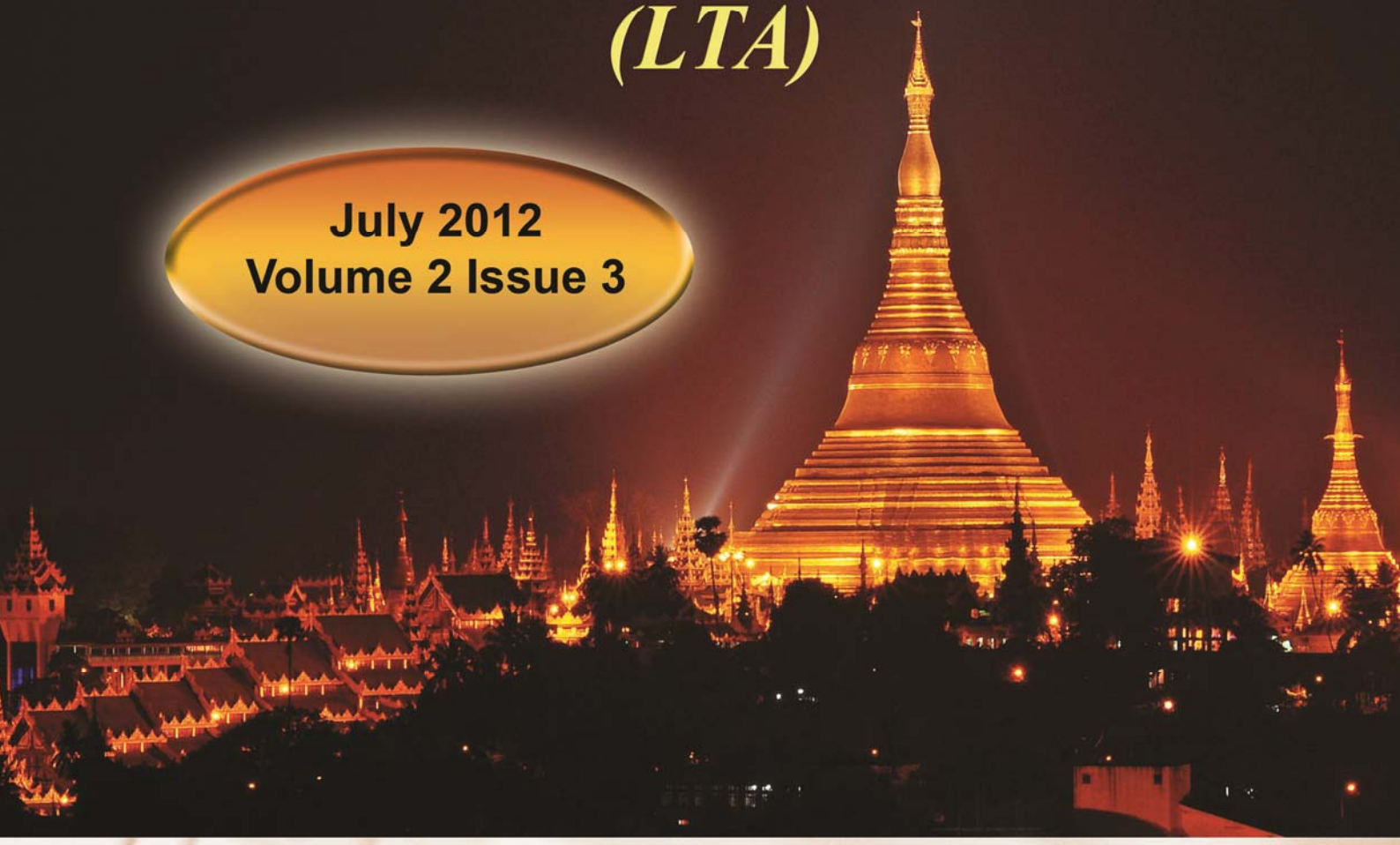
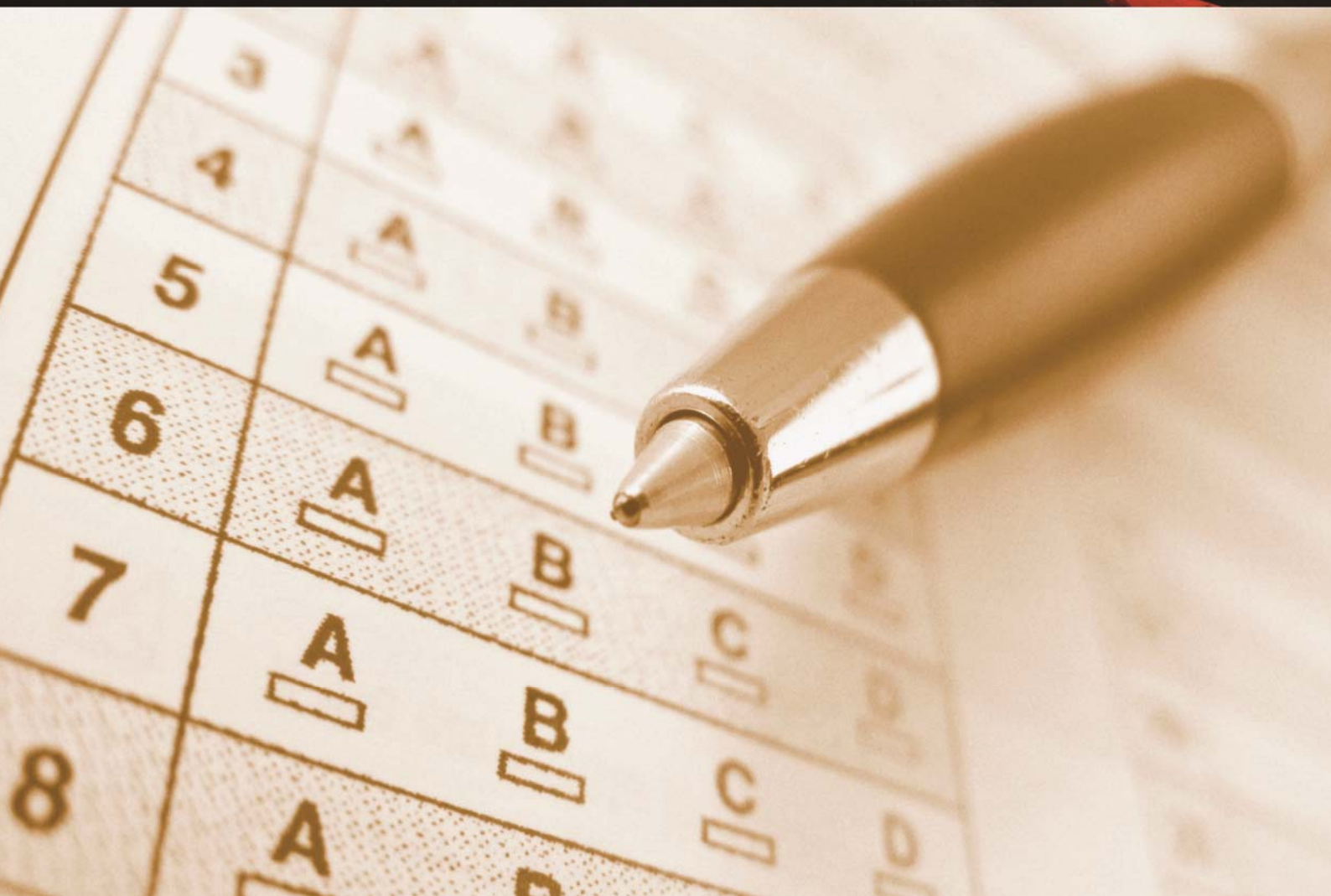
# *Language Testing in Asia (LTA)*

July 2012
Volume 2 Issue 3





**LANGUAGE TESTING IN ASIA (LTA)**

www.languagetestingasia.com/
Information on this title: http://www.languagetestingasia.com/



Editor: Dr Robert Kirkpatrick
Production Editor: Hmone Lian Hlaing

ISSN 2229-0443

## Editorial Board Members

# Contents

# Cognates in Vocabulary Size Testing – a Distorting Influence?

**EOIN JORDAN**
*Xi'an Jiaotong-Liverpool University*

**Bio Data:**
Eoin Jordan (MA, University of Nottingham) currently works as an English tutor at Xi'an Jiaotong-Liverpool University in China. His research interests include vocabulary acquisition, L1 transfer, and English pronunciation.

**Abstract**
This article examines the issue of cognates in frequency-based vocabulary size testing. Data from a pilot study for a cognate-controlled English vocabulary size test was used to assess whether a group of Japanese university English learners (n=60) were more successful at responding to cognate items than noncognate ones in three 1000 word frequency bands on a Japanese-English translation task. The results showed a statistically significant difference between scores achieved on cognate and noncognate items at the 2000 and 3000 frequency levels, but not at the 1000 frequency level. The findings suggest that cognate items may be easier for test-takers to respond to than noncognate ones of similar frequency, indicating the importance of ensuring that their respective proportions in tests are representative of those inherent in the frequency bands they have been sampled from. It is also argued that such representativeness may best be achieved via a stratified item sampling approach.

*Keywords:* vocabulary, test, cognates, loan-words, Japanese, English

## Introduction

Cognate words, that is those that '[come] naturally from the same root, or [represent] the same original word, with differences due to subsequent separate phonetic development' (OED Online, 2010), are undoubtedly encountered by learners from many L1 backgrounds during their study of English. While it is commonly known that a large number of English words share a root with other European tongues, it is also the case that some more linguistically distant languages have absorbed a great deal of vocabulary from, or that is cognate with, English too. In particular, a body of work by Daulton (1998, 1999, 2003, 2008) suggests that a large proportion of high frequency English vocabulary is cognate for Japanese learners, and asserts that these words can be utilized to assist Japanese students in their language studies. It should be noted that the vocabulary items which Daulton examines are also sometimes described as 'loan words' (as in Kay (1995)). While there is some controversy over

which term is most appropriate, these items are referred to as cognates here in order to maintain consistency with Daulton's work.

The effect that cognateness of a word has on how easy it is to recognize and learn has been investigated in depth by several researchers (De Groot & Keijzer, 2000; Ellis & Beaton, 1993; Hall, 2002; Lotto & De Groot, 1998). Their findings indicate that cognate words are indeed easier to acquire than noncognate ones, which in turn suggests that cognates do have the potential to be utilized effectively in language learning. This viewpoint is endorsed by Nation (2003), who argues that noticing of cognates is a valuable vocabulary expansion strategy.

Despite attention being paid to the effect of cognates on second language learning, there has been relatively little research conducted on the effect that they might have in language testing (although see Meara, Lightbown and Halter (1994) regarding French-English cognates in vocabulary testing). In fact, to the best of the author's knowledge no work has been done at all on this issue in the Japanese context. This is surprising, given that the presence of cognate vocabulary has previously been highlighted as an important problem in some well-known vocabulary tests such as the Vocabulary Levels Test (VLT) (Nation, 1990; Read, 2000) and Yes/No format tests (Eyckmans, 2004). In particular, it is possible that cognates may have a distorting effect on results for frequency-based vocabulary tests where a small number of items are randomly sampled from a large frequency band of words. When the sampling rate is low (for example, 10 items taken from a 1000 word frequency band, as in the Nation's Vocabulary Size Test (VST) (Beglar, 2010)) there is a considerable chance that the proportion of cognate items in the test will not be representative of the proportion of cognates in the frequency band. If cognate items are easier for test-takers than noncognate ones of the same frequency level, then this has the potential to affect results.

This article examines data from an English-to-Japanese translation task that was used as a criterion measure for a cognate-controlled frequency-based English vocabulary size test. A stratified sampling approach to Japanese cognate words was employed in the instrument's construction; thus allowing for a comparison to be made between scores on cognate and noncognate items in several of its frequency levels. It was anticipated that participants would score higher on average for cognate items in each frequency band, which would suggest that these items were indeed easier for respondents.

## Method

### Participants

The data examined in this article was collected from a total of 61 participants at Ritsumeikan Asia Pacific University in Japan; 59 Japanese undergraduate students and two Japanese English tutors (some Chinese and Korean undergraduate students had also participated in the study, but their results were discarded as it was felt that they may have been influenced by Japanese language difficulties). 56 of the undergraduates were studying on a 'Fundamental English' course, which had a paper-based TOEFL target score of 450, while the remaining 3 were registered on the university's 'Intermediate English' course, which had a target score of 500. The

students participated in the study under the supervision of their tutors during allocated class time, although Intermediate students were also given the option of engaging in an alternative activity during their class, which was the likely cause of the low number of respondents at this level. The two Japanese English tutors had both completed Masters degrees at English language institutions in Britain and America, and they participated in the study in their own time. Participants were not offered any rewards and were required to fill out an online consent form, which was included as part of the research instrument.

**Instruments**

The instrument used for the study consisted of an online version of a translation task followed by the new vocabulary size test mentioned in the introduction, which had corresponding vocabulary. However, as it is only the data collected from the translation task that is of interest here, a full description of the vocabulary size test is beyond the scope of this article. The translation task consisted of 100 items, sampled to cover the first 5000 words of English, so 20 items represented each 1000 word frequency band. This was expected to be sufficient to measure the vocabulary sizes of Japanese university students, who made up the majority of the participants; previous estimates of Japanese university students' average vocabulary sizes were 2000 (Shillaw, 1995) and 2300 (Barrow Nakanishi & Ishino, 1999). Translation was chosen as a criterion measure in line with Nation (2001), Eyckmans (2004), and Laufer and Goldstein (2004), who all favour it as the most thorough method by which to test receptive vocabulary knowledge.

**Word Lists**

The translation task was constructed from filtered versions of Paul Nation's (2009) British National Corpus (BNC) word family lists. These are frequency-based lists of word families (Bauer & Nation, 1993), defined in line with Bauer and Nation's Level 6 criteria (Nation, personal communication). Nation's data was filtered using the JACET 8000 lists (Aizawa, Ishikawa & Murata, 2005), which claim to represent the most important 8000 English word families for Japanese students. These are partly based on the BNC, but are also reflective of English teaching materials in Japan. It was intended that this process would remove any 'outliers'; that is, words which were considerably more or less familiar to Japanese learners than Nation's frequency levels indicated.

Each of Nation's frequency banded lists of headwords was inputted into the JACET 8000 Level Marker (Shimizu, 2009). The results, illustrated in Table 1, showed which thousand word JACET 8000 band each of Nation's headwords was located in. A small group of words were found to place very differently on the two lists; one example of this was the word 'confer', which was ranked in the 5000 band of JACET 8000 despite being in Nation's 1k list.

Table 1
*Nation's BNC lists divided into JACET 8000 levels*

| JACET 8000 lists | Nation's BNC lists | | | | |
|---|---|---|---|---|---|
| | 1k | 2k | 3k | 4k | 5k |
| 1000 | 745 | 187 | 15 | 12 | 7 |
| 2000 | 162 | 418 | 144 | 52 | 8 |
| 3000 | 31 | 172 | 270 | 140 | 60 |
| 4000 | 24 | 127 | 148 | 154 | 77 |
| 5000 | 8 | 42 | 160 | 142 | 143 |
| 6000 | 3 | 23 | 118 | 136 | 137 |
| 7000 | 0 | 2 | 41 | 98 | 124 |
| 8000 | 0 | 1 | 15 | 74 | 102 |
| Other | 9 | 26 | 89 | 191 | 342 |
| Unidentified | 18 | 2 | 0 | 1 | 0 |
| Total | 1000 | 1000 | 1000 | 1000 | 1000 |

The translation task was to consist of 20 headwords from each 1000 word frequency band, which meant that a single item would be representative of 50 words. In line with this, JACET 8000 word bands which had an overlap of less than 50 with any of Nation's individual frequency levels were excluded from that particular level. If included, these words would have been an over-representation of that JACET 8000 band. The exclusions meant that there should have been fewer outliers in terms of word difficulty for Japanese learners of English within each frequency level.

**Stratified Sampling of Cognates**

In order to ensure that cognates in the word lists were sampled in as accurate proportions as possible, it was first necessary to determine which of the word families featured Japanese cognates as their members and which did not. A list of English cognates for Japanese English learners constructed by Daulton (2003) was used as a starting point for this assessment. Daulton's work was based on one of Paul Nation's previous lists of the most common 3000 word families of English, and he found that a surprisingly high proportion of these appeared to have Japanese loan word equivalents. He highlights that not all of these cognates are equally closely related to their English equivalents, but for the purposes of this study it was felt that the most meaningful division that could be made was between cognates and noncognates; this difference was likely to have more effect on learners' ability to identify a correct form-meaning link than the difference between any two levels of positive cognateness. However the use of Daulton's lists presented two problems: (i)

The lists only covered the initial 3000 word families of English, whereas the translation task in this study was to cover filtered lists of the first 5000 word families, (ii) The first 3000 word families on which Daulton's lists were based were different to Nation's more recent lists.

In response to the two problems highlighted above, it was decided to supplement Daulton's (2003) list with cognate words derived from a Japanese corpus frequency list, and then to ascertain the intersection between this combined list and the filtered lists described in section 2.2.2. . Time and resource constraints meant that it was not possible to carry out an thorough empirical investigation into which Japanese loan words that corresponded with English words on the filtered lists were known by Japanese university students, rather the procedures described here were intended to provide a rough estimate of the number of cognates, with the aim of contributing to the rigorousness of the study.

**Assessing which Japanese loan words are widely known among the Japanese population.** Just because a Japanese loan word equivalent exists for an English word, it does not follow that this loan word will be in general usage. Likewise, the recent abundant usage of *katakana* (the Japanese script in which most recent foreign loan words from European languages are written) means that even though an internet search may reveal the existence of a few instances of a *katakana* word, this does not mean that it will be known by the majority of the Japanese population. In fact, the Japanese government has recently made efforts to stem the flow of new, difficult to understand loan words into Japanese by proposing alternative phrasings that utilize *kanji* (Chinese characters) to express the same concepts (The National Institute for Japanese Language, 2006). Bearing these factors in mind, one way of identifying loan words that are likely to be well known is to search for them in the upper range of a corpus frequency list; if such words are used frequently then it is likely that they will be familiar to most Japanese speakers.

There are very few publicly available large balanced general Japanese corpora (Goto, 2003; Ueyama, 2006); however a frequency list of lemmas from Serge Sharoff's internet corpus of Japanese (Sharoff, 2009) was considered to be suitable for the purposes of identifying high frequency loan words. As the corpus was assembled through the internet, computer and internet related words such as *kurikku* ('click') and *netowaaku* ('network') were ranked a lot higher than would be expected if it had covered spoken language. One can reasonably expect, however, that the majority of young adults studying English within the education system in Japan are at least to some extent familiar with the internet; thus it seemed plausible to argue that high frequency words on this list were generally well known among the population of young adult English learners.

Loan words in Japanese are usually written in the *katakana* script (Tohsaku, 1993). With this in mind, all of the *katakana* words from the first 10,000 entries of the lemma list were extracted for further analysis. The lemma frequency list actually contained 15,000 entries; however in order to be as confident as possible that the words extracted would definitely be known, only the first 10,000 lemmas were analyzed. Although no clear guidelines could be found on how many lemmas adult native speakers of Japanese are likely to know, the Japanese Language Proficiency

Test (The Japan Foundation, 2008) states that the vocabulary attainment target for the highest level test (Level 1) is knowledge of 10,000 words, and that this level is sufficient to operate in all areas of everyday life. It is not clear whether this refers to individual words, lemmas or word families, but it was felt that extracting *katakana* words from the first 10,000 lemmas of the frequency list would at least provide a conservative estimate of loan words in general usage.

The list of *katakana* words obtained from the corpus frequency list was translated through two online translation sites (Breen, 2010; Google Translate, 2009) and the English translations and *katakana* words checked for consistency. Although a rather subjective measure, any words that had been given English translations which the author felt did not make sense were also checked in the Wisdom Japanese-English Dictionary (Onishi, 2008). This is a widely used dictionary that lists definitions according to frequency of occurrence in an English corpus. Words that were found not to have English cognates were removed at this stage. The list of English translated words was then combined with the list of Daulton's words and put through the Range program (Heatley, Nation & Coxhead, 2002) to check for correspondence with the filtered versions of Nation's 1k to 5k lists on which the translation task would be based. If a word family contained any words that were cognates then that family was classified as cognate, on the grounds that if knowledge about one member could be inferred then this should then allow learners to infer knowledge about the other related members. Table 1 illustrates the numbers of word families designated as cognates.

**Calculation of cognate sampling ratios.** For reasons relating to the structure of items on the test that the translation task was used as a criterion measure for, sampling ratios of cognates and noncognates for each frequency band were calculated out of 10 rather than out of 20. The resultant ratios (displayed in Table 2) were multiplied by two, and then used to determine the numbers of different item types selected at random from the cognate and noncognate filtered word frequency lists for inclusion in the translation task.

Table 2
*Breakdown of Cognates and Noncognates in the Filtered Frequency Lists and TFVST*

| List | Cognate total | Noncognate total | Translation task cognates (/10) | Translation task noncognates (/10) |
|---|---|---|---|---|
| Filtered 1k | 572 | 335 | 6 | 4 |
| Filtered 2k | 466 | 438 | 5 | 5 |
| Filtered 3k | 267 | 573 | 3 | 7 |
| Filtered 4k | 150 | 837 | 2 | 8 |
| Filtered 5k | 64 | 921 | 1 | 9 |

**Presentation**
The translation task was divided into four sections of 25 items each in order to match with the format of the vocabulary size test. On each page of the instrument, 25 English words were presented with blank response boxes next to them. Japanese text at the top of the page instructed participants to fill in the most appropriate Japanese translation for each English word into its neighbouring response box. The instructions also stated that they should enter an 'X' in the relevant box if they did not know a translation for a word. After completing each section, participants were required to click on a button at the bottom of the screen that would move them on to the next section. They were not able to advance onto the next section until they had entered something into every response box on their current section. The instructions at the start of the test also informed students that they would not be able to return to completed sections of the test.

**Procedures**
Students were directed to the instrument via a temporary link using the university's online learning system. The two teachers who helped with administration of the study were briefed on the instrument in advance and gave a short explanation about it to their classes. Students could ask questions if they wanted to and were supervised throughout the whole procedure. The two Japanese English teacher participants were contacted individually. They agreed to participate during their free time and were sent a link to the instrument by email. All participants were told to follow the instructions presented on the screen during the study, which explained what was required of them in full.

**Data Analysis**
The data was downloaded and translations were then marked in accordance with the lenient marking scheme used by Eyckmans (2004). This meant that levels 1, 2 and 3 of her taxonomy (shown in Figure 1) were accepted as correct.

*Figure 1.* Marking taxonomy used for the translation task (from Eyckmans, 2004 p.81)

1. Correct translation
2. Correct translation but wrongly spelled or typed
3. Mistakes due to grammatical category
4. Undoubtedly incorrect translation or no response (X)

A lenient marking scheme was felt to be appropriate as the aim was to evaluate whether students had some level of knowledge of the form-meaning link, not to assess grammatical knowledge or Japanese language ability. Translations were allowed if they could be found in either Aizawa et al. (2005) or the Wisdom English-Japanese Dictionary (Inoue & Akano, 2008), or if they were judged to have the same meaning as translations in these sources and were listed in the online ALC database (SPACEALC, 2000). Correct responses were awarded 1 mark, incorrect

responses 0. Items that had not been categorized as cognates during test construction for which a *katakana* loan word was marked as a correct answer were again noted down, then later excluded from comparison analyses between responses for cognate and noncognate items. There were 17 of these items in total, the majority occurring in the lower frequency levels, suggesting that the cognate lists constructed for this test are probably rather conservative estimates of the total number of cognates within the filtered word lists.

## Results and Discussion

### General Results

Descriptive statistics and the Cronbach Alpha reliability coefficient for the translation task are displayed in Table 3. The Alpha coefficient was sufficiently high, while the mean score and standard deviation suggested that the frequency range of the task was appropriate for the participants. It was also noted that scores on the translation task decreased on average with each of the five frequency bands (see Table 4), suggesting, as expected, that participants were less familiar with lower frequency vocabulary. In this respect the task had performed in a similar manner to other tests of vocabulary size.

Table 3
*Descriptive statistics and reliability coefficient for the translation task*

| *N* = 60 | *M* (/100) | *SD* | Reliability (Cronbach α) |
|---|---|---|---|
| Translation task | 45.28 | 12.95 | .94 |

Table 4
*Word frequency level comparisons of scores from the translation task*

| N=60 | ***M* (*SD*)** |
|---|---|
| | Translation task |
| 1000 frequency level (/20) | 17.13 (2.40) |
| 2000 frequency level (/20) | 9.90 (3.43) |
| 3000 frequency level (/20) | 6.73 (3.24) |
| 4000 frequency level (/20) | 5.93 (3.41) |

| | |
|---|---|
| 5000 frequency level (/20) | 5.83 (2.73) |

**Differences in Performance on Cognate and Noncognate Items**

Given the small numbers of cognate items available to be analysed at the 4000 and 5000 frequency levels, comparisons between the performance of cognate and noncognate items were restricted to the first three frequency bands (although it was noted that average scores on the few cognate items in the lower frequency bands were a great deal higher than those for equivalent noncognate items). Table 5 presents a comparison of correct response rates for cognate and noncognate items in each of these frequency levels, which shows that the correct response rate for cognate items was slightly lower at the 1000 frequency level, but considerably higher at the 2000 and 3000 frequency levels.

Table 5

*Comparison of correct response rates for cognate and noncognate items at different frequency levels*

| | ***M* (*SD*, *n*)** | |
|---|---|---|
| | Cognate items | Noncognate items |
| 1000 frequency level | .84 (.13, 12) | .87 (.14, 6) |
| 2000 frequency level | .73 (.18, 10) | .26 (.21, 10) |
| 3000 frequency level | .76 (.22, 6) | .14 (.19, 10) |

To investigate further, a two-way ANOVA was conducted to examine the effects of word frequency level and item type (cognate or noncognate) on correct response rates. Levene's test for equality of error variances suggested that there was homogeneity of variance between groups; thus making the two-way ANOVA viable technique for this analysis. A significant interaction was found between the two effects, $F$ (2, 48) = 15.226, $p$ = .000. Simple main effects analysis showed that there were significant differences between correct response rates on cognate and noncognate items at the 2000 frequency level ($p$ = .000) and 3000 frequency level ($p$ = .000), but not at the 1000 frequency level ($p$ = .737). The lack of a statistically significant difference at the 1000 frequency level may be explained by students' very high average scores in this section, which suggested that most participants were familiar with the vast majority of vocabulary at this level, regardless of whether it was cognate or not. In general, however, the results justified the stratified sampling of cognate items in the test; cognate items appeared to be easier for participants to

answer correctly, and if random sampling had been employed then this may have resulted in unrepresentative proportions of cognates in frequency levels.

The implications of these results are that frequency-based L2 vocabulary size tests for homogeneous L1 contexts, particularly those that have a low sampling rate, run a considerable risk of producing distorted results if they fail to take account of cognate words during item selection. The stratified sampling procedure outlined in this report is one method that can be used to improve the situation by sampling cognates more proportionately; however, this is reliant on the presence of accurate data about cognates, which may not exist for many language pairs. Indeed, the number of valid *katakana* translations provided for supposedly noncognate words in the translation task reported on here suggests that even the lists of Japanese-English cognates used in this study can only be considered as conservative estimates at best. Accordingly, an important focus for future research in this area will need to be the collection of more accurate data on cognates between widely used languages.

**Limitations**

This report describes a study that was conducted on one small group of Japanese university students and teachers, many of whom had a similar general English level. The effect that a wider and more evenly spread range of English proficiencies among participants would have had on results is not clear; thus the findings here are somewhat limited in terms of their generalizability. A lack of incentive to perform well on the tests may also have resulted in participants not engaging with the material in an earnest manner. Finally, the number of cognate items in the lower frequency bands was, out of necessity, very small, which meant that other properties of these words may have had some influence on results here.

## Conclusion

In general mean scores on cognate items were significantly higher than mean scores on noncognate items in the same frequency level. This finding appeared to justify the cognate-controlled design of the test, suggesting that a stratified sampling approach to cognates is likely to produce more accurate estimates of vocabulary size than random sampling of items from frequency bands (particularly when the sampling rate is low). It is the hope of the author that other similar studies are conducted in the future to add further weight to this assertion. The major problem in implementing stratified sampling of cognates in vocabulary size tests for homogenous L1 groups is likely to be a lack of reliable data on which words are, in fact, cognate with L2. The analysis of the data produced in this study suggested that the lists for English and Japanese used here did not cover all of the cognates, particularly at lower frequency levels, and for many other language pairs reliable data may not exist at all. Accordingly, one other area in which further research is needed is the development, using both empirical investigation and corpus data, of reliable lists of cognates for different language pairs.

**References**

Aizawa, K., Ishikawa, S., & Murata, T. (2005). *JACET 8000 English Basic Words*. Tokyo: Kirihara-shoten.

Barrow, J., Nakanishi, Y., & Ishino, H. (1999). Assessing Japanese college students' vocabulary knowledge with a self-checking familiarity survey. *System*, *27*(2), 223-247.

Bauer, L., & Nation, I.S.P. (1993). Word families. *International Journal of Lexicography, 6*, 253-279.

Beglar, D. (2010). A Rasch-based validation of the vocabulary size test. *Language Testing, 27*, 101-118.

Breen, J.W. (2010). *WWWJDIC: Online Japanese dictionary service*. Retrieved February 17, 2010 from http://www.csse.monash.edu.au/~jwb/cgi-bin/wwwjdic.cgi?9T

Daulton, F.E. (1998). Japanese loanword cognates and the acquisition of English vocabulary. *The Language Teacher, 22*(1), 17-25.

Daulton, F.E. (1999). List of high-frequency baseword vocabulary for Japanese EFL students. *The Internet TESL Journal, 5*(11). Retrieved February 17, 2010 from http://iteslj.org/lists/Daulton-BasewordVocabulary.html

Daulton, F.E. (2003). List of high-frequency baseword vocabulary for Japanese EFL students #2. *The Internet TESL Journal, 9*(3). Retrieved February 17, 2010 from http://iteslj.org/lists/Daulton-BasewordVocabulary.html

Daulton, F.E. (2008). *Japan's built-in lexicon of English-based loanwords*. Clevendon: Multilingual Matters.

DeGroot, A. M. B., & Keijzer, R. (2000). What is hard to learn is easy to forget: the roles of word concreteness cognate status, and word frequency in foreign-language learning and forgetting. *Language Learning*, 50, 1–56.

Ellis, N.C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, 43, 559-617.

Eyckmans, J. (2004). *Measuring receptive vocabulary size*. LOT (Landelijke Onderzoekschool Taalwetenschap).

Goto, H. (2003). *言語理論と言語資料：コーパスとコーパス以外のデータ*[Linguistic Theories & Linguistic Resources: Corpora and Other Data]. *日本語学* [Japanese Language Studies] *22*, 6-15. Retrieved February 17, 2010 from http://www.sal.tohoku.ac.jp/~gothit/nhnggk0304.html

Hall, C.J. (2002). The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development. *IRAL - International Review of Applied Linguistics in Language Teaching*, *40*(2), 69-87.

Google Translate (2009). *Google translate*. Retrieved February 17, 2010 from http://translate.google.com

Heatley, A., Nation, I.S.P., & Coxhead, A. (2002). *RANGE and FREQUENCY programs*. Retrieved February 17, 2010 from http://www.vuw.ac.nz/lals/staff/Paul_Nation

Inoue, N., & Akano, I. (2008). *The Wisdom English-Japanese dictionary*. Tokyo: Sanseido.

Kay, G. (1995). English loanwords in Japanese. *World Englishes, 14*(1), 67-76.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning, 54*(3), 399-436.

Lotto, L., & De Groot, A.M.B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning, 48*, 31–69.

Meara, P., Lightbown, P.M., & Halter, R.H. (1994). The effect of cognates on the applicability of YES/NO vocabulary tests. *Canadian Modern Language Review, 50*(2), 296-311.

Nation, I.S.P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I.S.P. (2003). The role of the first language in foreign language learning. *Asian EFL Journal, 5*(2), Retrieved February 17, 2010 from http://www.asian-efl-journal.com/june_2003_PN.php

Nation, I.S.P. (2009). *BNC 14k lists*. Retrieved February 17, 2010 from http://www.lextutor.ca/vp/bnc/nation_14/

OED Online (2010). *Oxford English Dictionary*. Oxford University Press. Retrieved January 3, 2011 from http://www.oed.com/view/Entry/35870?redirectedFrom=cognate

Onishi, T. (2008). *The Wisdom Japanese-English dictionary*. Tokyo: Sanseido.

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

Sharoff, S. (2009). *Large corpora used in CTS*. Retrieved February 17, 2010 from http://corpus.leeds.ac.uk/list.html

Shillaw, J. (1995). Using a word list as a focus for vocabulary learning. *The Language Teacher, 19*(2), 58-59.

Shimizu, S. (2009). *JACET 8000 level marker*. Retrieved December 15, 2009 from http://www01.tcp-ip.or.jp/~shin/J8LevelMarker/j8lm.cgi

SPACEALC (2000). *Eitarou on the web*. Retrieved February 17, 2010 from http://www.alc.co.jp/

The Japan Foundation (2008*). The Japanese-language proficiency test*. Retrieved February 17, 2010 from http://www.jlpt.jp/e/

Tohsaku, Y. (1993). *Yookoso!: Invitation to contemporary Japanese*. New York: McGraw-Hill.

Ueyama, M. (2006). Evaluation of Japanese Web-based Reference Corpora: Effects of Seed Selection and Time Interval. In Baroni, M., & Bernardini, S. (Eds.), *Wacky! Working papers on the web as corpus*. Bologna: GEDIT. Retrieved February 17, 2010 from http://wackybook.sslmit.unibo.it/wackycontents.html

## Appendix

List of words used in the translation task

**SECTION A**
ADVANCE, CHARACTER, DRY, EASY, FARM, GERMANY, GROW, HAND, HERE, INDUSTRY, INFORM, MAJOR, NEW, PARTY, PROCESS, PROMOTE, PURE, SIT, SITE, SOCIAL, SUGGEST, THEN, TIME, WINE, WISE.

**SECTION B**
BIBLE, CASUALTY, CHAPEL, CONTRIBUTE, CRITERION, CUSHION, DIMENSION, FETCH, FORMAL, HILL, INTELLIGENCE, JEANS, LIBERAL, MANUAL, NEVERTHELESS, ORCHESTRA, PUBLISH, REPLY, RIVER, SKY, SPIN, TREMENDOUS, TRIVIAL, ULTIMATE, VEGETABLE.

**SECTION C**
ANTIQUE, ARTIFICIAL, BAIL, BEE, BUBBLE, CEASE, DELIBERATE, EMPIRE, FIN, GALLON, IRONY, ISRAEL, IVY, MUTUAL, OUTRAGE, PALM, PUNISH, RAVE, RECITE, REVEAL, SHATTER, SOAP, TRACTOR, VERIFY, WHEELCHAIR.

**SECTION D**
ADMINISTER, BLEND, BLOUSE, BROOM, DEADLINE, DEFICIENCY, DIVINE, EMIGRATE, FLUID, FOG, GASP, INTERCEPT, JAZZ, JURY, MINIATURE, MULTITUDE, PHYSICS, PIERCE, PUBLICIZE, RENDER, RETREAT, SNAKE, TESTAMENT, TRADESMAN, UNDERWEAR.

# An Investigation into the Iranian EFL Language Learners' Attitudes on TOEFL iBT

**KAMAL HEIDARI SOURESHJANI, PARISA RIAHIPOUR & AREZOO SAFIKHANI**
*Islamic Azad University, Shahrekord Branch, Iran*

**Bio Data:**
Kamal Heidari Soureshjani holds M.A in TEFL from Shiraz University and is a Young Researchers Club Member. He taught English courses and IELTS at different institutes in Shiraz and is presently the academic member of Azad University, Shahrekord branch.

Parisa Riahipour has M.A. in TEFL and is now an instructor in Islamic Azad university, Shahrekord Branch. She has been teaching English in different institutes and in different levels for about three years.

Arezoo Safikhani also has M.A. in TEFL and has been teaching English in different institutes and in different levels for several years. She has compiled a set of articles published in different journals.

**Abstract**
The present study, which was probably the first of its kind in Iran, aimed at investigating the attitudes of Iranian iBT candidates towards the iBT test. It especially addressed the role of gender in the attitude of the participants of the study on iBT test. Accordingly, an attitude questionnaire was developed based on the theoretical framework of the study and then administered among Iranian iBT candidates who had taken the test in Shiraz, Iran. The collected data was then analyzed through chi-square and T-test to see if there was any meaningful relationship between candidates' sex and their attitudes towards the iBT test. The study revealed that first, most of the participants in the study had a positive attitude toward iBT and second, it was found out that the candidates' gender had no significant role in their attitudes towards iBT test. Implications can be drawn for all the stakeholders including candidates intending to sit for the test, institutes running iBT preparation program and teachers wishing to teach such programs.

*Key terms:* iBT test, t-test, Iranian candidates' attitudes

## Introduction

The *Test of English as a Foreign Language* Internet-Based test (TOEFL iBT) was designed as a measure of English ability for university academic studies in North America. It was introduced in September 2005 and gradually spread worldwide during 2005 and 2006. TOEFL iBT was developed in response to a request by

institutions that would measure non-native speakers' ability to communicate in English in an academic setting. iBT which is a 100% academically-focused test, meaning that it measures the kind of English language used in academic settings, specifically has been designed to measure the ability to communicate by combining, or integrating, all four language skills.

The concept of attitude has been the focus of attention in explanation of human behavior offered by social psychologists. Attitude is usually defined as a disposition or tendency to respond positively or negatively towards a certain thing such as an idea, object, person or situation. Students may have positive or negative attitudes towards the language they want to learn or the people who speak it. Having positive attitudes is claimed to be one of the reasons which make students perform better.

## Literature Review

### History of TOEFL

The Test of English as a Foreign Language, better known as TOEFL, is designated to measure the English language proficiency of people whose native language is not English. TOEFL scores are accepted by more than 6000 colleges, universities, and licensing agencies in 130 countries. The test is also used by governments, and scholarship and exchange program worldwide. A list of institutions and agencies that accept TOEFL scores is available on the TOEFL website at www.ets.org/toefl.

A national council on the testing of English as a foreign language was formed in 1962; its members were representatives of more than 30 private organizations and government agencies concerned with the English language proficiency of nonnative speakers of English who wished to study at colleges and universities in the United States. The council supported the development of the TOEFL test for use starting in 1963-64. Financed by grants from the Ford and Dan Forth Foundations, the TOEFL program was first administered by the Modern Language Association. In 1965, the College Board and Educational Testing Service (ETS) assumed joint responsibility for the program. Because many who take the TOEFL test are potential graduate students, a cooperative arrangement for the operation of the test was entered into by ETS, The College Board, and the Graduate Record Examinations Board in 1973. Under this arrangement, ETS is responsible for administering the TOEFL program with guidance from the TOEFL Board.

The test originally contained five sections. As a result of extensive research, a three-section test was developed and introduced in 1976. In July 1995, the test item format was modified somewhat within the same three-section structure. In recent years, various constituencies called for a new TOEFL test that 1) be more reflective of communicative competence models; 2) include more constructed-response tasks and direct measures of writing and speaking; 3) include tasks that integrate the language modalities tested; and 4) provide more information than the paper-based TOEFL test (TOEFL PBT) about the ability of international students to use English in an academic environment. Accordingly, the TOEFL Board initiated a board effort under which language testing will evolve in the twenty-first century. The introduction of the computer-based TOEFL test (in TOEFL CBT) 1998 was the first incremental step in this broad test-improvement effort.

The next step was the introduction of an Internet-based TOEFL test (TOEFL iBT) in September 2005. The test was first launched in the United States, and gradually rolled out worldwide during 2005 and 2006. TOEFL iBT assesses all four language skills (reading, listening, speaking and writing) that are important for effective communication. TOEFL iBT emphasized integrated skills and provides better information to institutions about students' ability to communicate in an academic setting and their readiness for academic coursework.

As TOEFL iBT was introduced in an area, TOEFL CBT was discontinued after a period of overlap to insure a smooth transition to TOEFL iBT. The final administration of TOEFL CBT was held in September 2006. TOEFL PBT will continue to be offered on a limited basis to support the TOEFL testing network in areas where TOEFL iBT is not available.

**Attitude**

The concept of attitude has been the focus of attention in explanation of human behavior offered by social psychologists. Johnson & Johnson (1998) define attitudes as opinions, beliefs, ways of responding, with respect to some sets of problems. So, they contain or closely relate to our opinion, belief and are based on our experiences. Attitudes represent a major connection between cognitive and social psychology since they are often related to interaction with others. They are strongly connected to feelings. According to Lange & James (1972), attitudes suggest a feeling for or against something. Ajzan (1988) believes that they are latent hypothetical characteristics that can be inferred from external & observable cues. Sarnoff (1970) has also defined attitude as a disposition to react favorably or unfavorably to an object, situation, person or event.

Attitudes, like other aspects of the development of cognitive & affective factors in human being, develop early in childhood and are the result of parents' & peers' attitudes of contact with people who are different in any number of ways and of interacting affective factors in human experience. Thus attitudes form a part of one's perception of self of others, and of the culture in which they are living (Brown, 2000).

**Attitudes toward Language, Language Learning and Language Tests**

Gardner and Lambert (1972) believe that attitudes toward language are defined in terms of different orientations towards language learning. Major orientations are called integrative and instrumental orientations. According to Oxford & Shearin (1994) attitude is one of the factors impacting motivation in language learning.

Attitudes that have been explored in relation to language learning range from anxiety about language and the learning situation, through attitudes to speakers of the second language, the country in which it is spoken, the classroom, the teacher, other learners, the nature of language learning, particular elements in the learning activities, tests and beliefs about learning in general (Johnson & Johnson, 1998). Chastain (1998) regards attitude as one of the variables contributing to second language learning. Besides, attitudes can be positive/negative. Bachmann (1976) has argued that high achievement causes positive attitudes and high motivation.

Mantle (1995) examined the language and culture attitudes of middle schools participating in a foreign language exploratory program. Results clearly revealed that many students enter their first language class with misconception about language learning that may hinder their progress or persistence in language study.

A host of studies have also investigated the relationship between attitudes and proficiency levels. Gardner (1985) as an example believes that attitudes and other affective variables are as important as aptitude for language achievement. A study carried out by Krajewska (1997) also showed a positive relationship between attitude & language abilities. Malallaha (2000) investigated the attitudes of Arab learners toward English and discovered that they have positive attitudes toward the English language and their proficiency in tests was positively related to their positive attitudes toward English.

On the whole, one can claim that having positive/negative attitudes towards a certain language can exert considerable effect on the learners' performance on a language test. By the same token, learners' attitudes toward a certain language proficiency test may affect their performance on their test. iBT candidates' attitudes towards iBT, therefore , might affect their overall score they get in this standardized test.

**Significance of the Study**

To the best of researchers' knowledge, no similar study has been done in the EFL context of Iran. In this perspective, therefore, the present study hopes to gain significance as the results can help all iBT –TOEFL stakeholders including candidates intending to participate in the test, institutes holding preparation program for iBT, teachers wishing to teach such programs and finally iBT test administrators who are running the test in Iran. The iBT venues must be undergone quality control processes more often and iBT administrators, ushers and examiners must be trained regularly so that everything goes well and in a standard fashion.

**Objectives of the Study**

The main objective of this study is to determine the attitudes of Iranian candidates towards the TOEFL iBT test. In addition, factors such as test environment, test rubric, candidates' age and sex and their relationship with candidates' attitudes will be investigated. The study therefore, seeks answers to the following questions:

1) What do Iranian candidates think of the TOEFL iBT test?
2) Does the gender of participants make a distinction in their attitudes?

## Method

**Participants**

The study was conducted with the participation of two independent groups of participants. The first group consisted of 150 Iranian iBT candidates who took the actual test in Shiraz University. All of them were from Iran and speak Farsi as their first language. They were 69 male and 81 female candidates ranging from 22 to 54 in age. They were also from different educational backgrounds. As to the second group of participants, 25 of the candidates apart from the first group, were randomly

picked out and interviewed. They were also both male and female (10 male and 15 female) and aged from 24 to 46 with different educational backgrounds.

**Instruments**

A questionnaire and an interview were employed in the study to gather the required data from the participants. As to the questionnaire, it was an attitude questionnaire (see appendix 1) developed by addressing the overall attitudes of the candidates towards iBT test based on the theoretical framework adopted for the study. The main purpose of the questionnaire distribution among the participants was to address the overall attitude of the candidate towards iBT and their attitudes towards the four component of the iBT test, i.e. listening, reading, writing and speaking. Finally, its reliability was estimated through Cronbach Alpha which yielded 0.75 and its validity was determined through being examined and confirmed by some related professors.

After taking the test by the participants, 25 of the candidates were randomly selected for the second instrument of the study that is, a semi-structured interview (See Appendix 2). The items of the questionnaire were in keeping with the items of the attitude questionnaire so that, the participants' responses would be more consistent. The point with regard to the participants who took the interview is that they were apart from the ones who filled in the attitude questionnaire.

**Data Collection**

In order to collect the data, one of the researchers of the study first distributed the questionnaire among the candidates and asked them to fill it in carefully. They were seated at a previously-prepared room with enough time for the filling out the forms.

As to the interview, the other researcher of the study asked some of the participants to do an interview about the test. From among about 50 candidates, just 25 of them agreed to be interviewed. Again, for the interview there was previously prepared a suitable and comfortable place so that the interviewees were able to feel relaxed and answer the questions in an ideal condition.

**Data Analysis**

To analyze the gathered data, two different methods of analysis were used. As to the questionnaire, in order to get knowledge with regard to the participants' perspectives about the iBT test, a chi-square was run. Then in order to ascertain whether or not the gender of participants has any impact on their attitudes a t-test was run. With regard to the interview results, the responses of the participants were described, analyzed, and compared with the questionnaire results.

## Results and Conclusion

Having conducted the study and gathered the required data, now in this section of the study the main results of the study are presented. To do so, the research questions of the study are presented one by one and then their answers based on the study findings are mentioned.

1) What do Iranian candidates think of the TOEFL iBT test?

The first research question of the study seeks the attitude of the candidates apropos of the iBT test. Table 1 which represents the attitude of the candidates in terms of

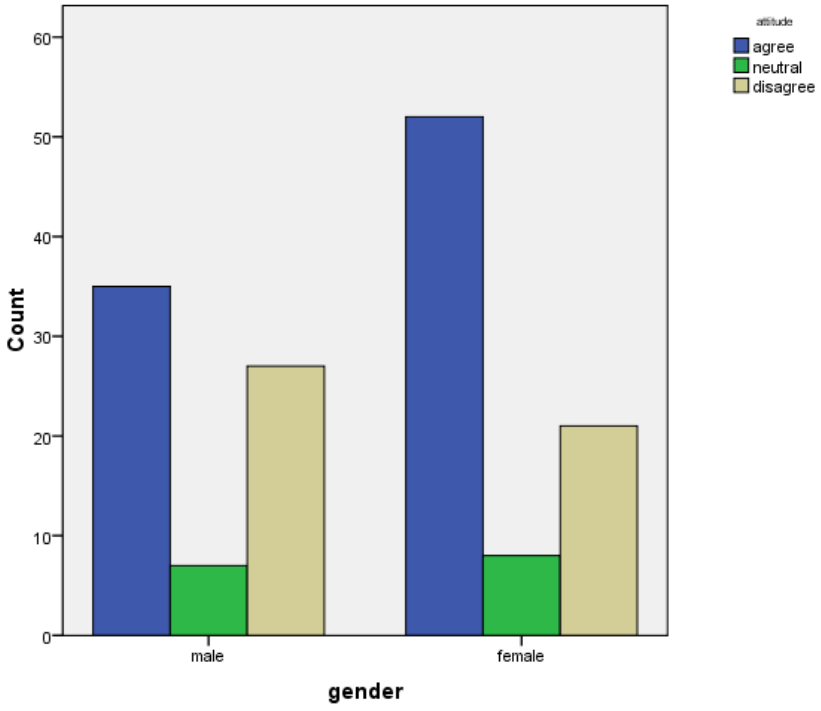
their frequency of responses to the questionnaire items. As the table clearly shows with regard to the male candidates they selected the choice of "agree" for most of the items of the questionnaire and in contrast, the choice of "neutral" was selected by them for the least frequency. As to the females, like the male ones, they have selected the choices "agree" and "neutral" for the most and least, respectively.

Table 1
*Gender and Attitude Frequency*

| | | attitude | | | |
|---|---|---|---|---|---|
| | count | agree | neutral | disagree | Total |
| gender | male | 1767 | 105 | 543 | 2415 |
| | female | 4232 | 2290 | 2388 | 8910 |
| Total | | 5999 | 2493 | 2833 | 11325 |

The same information presented in table 1 can be more vividly observed from figure 1 presented below. As it is self-evident, both male and females' agreed mostly with majority of the questionnaire items. In contrast, they chose "neutral" the least.

*Figure 1.* Gender and Attitude Frequency




The same information can be found in table 2 in the percentage form. This table, in comparison to the previous table, provides more tangible information as to

the candidates' responses. The table, in keeping with table 1, manifests that both males and females responded positively to the items of the questionnaire.

Table 2
*Gender and Attitude Crosstabulation*

| | | | attitude | | | |
|---|---|---|---|---|---|---|
| | | | agree | neutral | disagree | Total |
| gender | male | Count | 22 | 33 | 14 | 69 |
| | | % within gender | 31.9% | 47.8% | 20.3% | 100.0% |
| | | % within attitude | 60.4% | 46.5% | 37.8% | 46.0% |
| | | % of Total | 14.7% | 22.0% | 9.3% | 46.0% |
| | female | Count | 20 | 38 | 23 | 81 |
| | | % within gender | 24.7% | 46.9% | 28.4% | 100.0% |
| | | % within attitude | 62.6% | 53.5% | 47.2% | 54.0% |
| | | % of Total | 13.3% | 25.3% | 15.3% | 54.0% |
| Total | | Count | 42 | 71 | 37 | 150 |
| | | % within gender | 28.0% | 47.3% | 24.7% | 100.0% |
| | | % within attitude | 100.0% | 100.0% | 100.0% | 100.0% |
| | | % of Total | 28.0% | 47.3% | 24.7% | 100.0% |

In addition, table 3 labeled as chi-square tests reveals that since the reported Sig. value is larger than .05 ($p = .43$) therefore, the male and female candidates do not differ from each other in terms of their attitudes towards iBT test.

Table 3
*Chi-Square Tests*

| | Value | df | Asymp. Sig. (2-sided) |
|---|---|---|---|
| Pearson Chi-Square | 1.687[a] | 2 | .430 |
| Likelihood Ratio | 1.698 | 2 | .428 |
| Linear-by-Linear Association | 1.648 | 1 | .199 |
| N of Valid Cases | 150 | | |

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 17.02.

As to the interview results pertaining to this research question, most of the interviewees thought positively with regard to the iBT test. As an example, one of them stated that:

> *I have previously taken different forms of TOEFL whether paper-based or computer-based versions. Now that I took the iBT form too, I personally think that this form of test was much better than the other two forms in different terms such as the procedure of test taking, the time allocation, and also the preparatory actions like its enrolment.*

This candidate has the advantage of taking all other types of TOEFL. Therefore, her responses can be a good judge. She said that one of the advantages of iBT over other forms of the test is the procedure of test taking. That is, instead of sitting for a long time in our seats waiting for the proctors to bring the papers (as in paper-based form) and also using traditional stationery such as pen, pencil, eraser, etc. in the iBT form, we don't need any pen and pencil (just for some times like taking a note), we can easily access the questions on the net. The point needs to be noticed here is that although this kind of test-taking may be easier and more comfortable than the traditional forms, there are, however, some people who may not be able to work with modern equipments. In line with this point, one of the interviewers asserted that:

> *I had really studied for the test. Before I took the test I was certain that I will get an excellent score on the test. But when I started taking the test, I didn't know how to work with the machine. I got so nervous as the test starts. I couldn't use my knowledge on the test just due to lack of familiarity with the computer.*

This statement easily points to the point that although technology-based test taking can be a great help in the more effective, practical, and efficient test taking, it can also be a disaster for some. That is, there were some candidates who had problems like lack of familiarity with the used technology, lack of sufficient knowledge with regard to the format of the test, etc. Another candidate who was interviewed remarked that:

> *I think personally that iBT test was, all in all, better and more standardized in comparison to the other formats. However, there are still some problems. As an*

*example, I believe that with regard to the reading section of the test, there wasn't a balance between the length of texts, the included items and the allocated time. To put clearly, the length of texts followed by its related questions was my main difficulty with the test. I couldn't answer most of the items just due to lack of sufficient time.*

Some of the interviewed candidates positively agreed that iBT test was by far more stress free than the previous formats. One of them, for instance, added that:

*IBT test was really more standard than the previous formats. The main reason for this claim is that it is less threatening for the candidates. To my own experience, when I sat for the test I was somehow nervous initially but immediately after starting the test I was downright relaxed. I didn't have to wait for the proctors to distribute the papers and to announce the instructions loudly. I worked on my own pace and I felt no pressure on me.*

2) Does the gender of participants make a distinction in their attitudes?

Regarding the second research question about the effect of gender on the candidates' attitudes towards the iBT test, the results of t-test signaled no significant result. In other words, as it is conspicuous from table 4, there is no significant relationship between the gender of candidates and their attitudes (Sig. = .20 > .05).

Table 4
*Independent Samples Test*

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 95% Confidence Interval of the Difference | |
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | Lower | Upper |
| attitude | Equal variances assumed | .049 | .82 | -1.2 | 148 | .20 | -.15 | .11 | -.38 | .081 |
| | Equal variances not assumed | | | -1.2 | 145.062 | .20 | -.15 | .11 | -.38 | .081 |

Although no significant difference was observed between the male and female candidate in their attitudes about iBT test, table 5 shows that, by exploring the Mean and SD values, the female candidates were more consistent and similar in their attitudes in comparison to the male candidates' attitudes.

Table 5

*Group Statistics of Gender and Attitudes*

| | gender | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| attitude | male | 69 | 1.8 | .73 | .08 |
| | female | 81 | 2.0 | .71 | .08 |

## Conclusion

The present study was, in fact, an attempt to investigate the Iranian candidates' attitudes towards iBT test. It also examined the role of gender in their attitudes. To put it another way, the paper tried to cast light on this point that whether or not male and female iBT candidates differ in their attitudes towards the test. To achieve these purposes, 150 male and female Iranian iBT test candidates were selected and then a developed questionnaire was given to them to fill it in. Besides, 25 more candidates were also interviewed to gather their views about the test.

The study results revealed that most of the Iranian candidates looked positively at the iBT test. It was also found out that the gender of candidates has no significant role in their attitudes toward the test. In other words, the study showed that male and female candidates don't differ significantly in terms of their views about iBT test.

Implications can be drawn for all iBT stakeholders, those who intend to sit for iBT, those who run iBT preparation programs, and also those who develop the questions and administer the iBT test. Those in charge of administration of iBT in Iran should pay proper attention to the demands of Iranian candidates while taking the test. Equipping all the iBT centers with some facilities such as headphones for the listening section of iBT seems to be essential. Participants of this study viewed physical conditions of iBT such as light, temperature and chairs highly effective on their performance on the test and thus their attitude towards iBT. Test makers can also benefit from the findings of this study. They might decide to modify the listening section in a way so that it will reduce the stress and confusion of candidates who are not familiar with the format of the test. They might also increase the time limit for the reading section or also truncate the length of passages. Finally, those in charge of running iBT preparation programs can also take advantage of this and similar studies in that based on studies carried out on the same issue, listening and reading are the most difficult sections of iBT test. They should provide candidates with many authentic texts and encourage them to read extensively outside the classroom.

The study, despite the author's efforts, suffers from a set of limitations. First of all, the number of participants of the study was small, and the study was also limited to a single context hence, generalizing the study results may not be done with an absolute certainty. Consequently, in order to reach much more reliable findings further studies should be done with more participants from different Iranian contexts. Secondly, there may be some problems with regard to the instruments of the study especially the questionnaire. Some more items may need to be added to the questionnaire.

Further studies can be carried out to investigate issues related to iBT in Iran. Studies can be done involving the iBT listening and reading sections. Further studies can be run on the use of iBT for professional purposes or for migration. Studies on test preparation practices and investigation of the cognitive processes of iBT test takers can be done as well. Finally, studies can also be conducted investigating the relationship between candidates' stress and its effect on their performance on the test.

**Reference**

Ajzan, I. (1988): *Attitudes, personality, & behavior.* Stratford: Open University Press.

Bachman, N. (1976). Two measures of affective factors as they relate to progress on adult second language learning. *Working Papers on Bilingualism, 10,* 100-122.

Brown, D. H. (2000). *Principles of language learning and teaching (4th ed.).* White Plains, NY: Pearson Education.

Chastain, K. (1998). *Developing second language skills (2nd ed.).* Chicago: Rand Macually Press.

Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivations in second language learning*. Rowley, MA: Newbury House Publishers.

Gardner, R. C. (1985). *Social psychology and second language learning.* London: Edward Arnold.

Johnson, K., & Johnson, H. (1998). *Encyclopedic dictionary of applied linguistics*. Oxford: Blackwell Publishers Ltd.

Krajewska, U. (1997). Second language learning and the attitudes about foreigners. *Jezki-obce-w-szkole, 41,* 407-411.

Lange, D. C., & James, C. J. (1972). *Foreign language education: Reappraised*. Shokie, Illinois: National textbook company.

Malallaha, S. (2000). English in Arabic environment: Current attitudes to English among Kuwait University Students. *International Journal of Bilingual Educations & Bilingualism, 3,* 19-43.

Mantle-Bromley, C. (1995). Positive attitudes and realistic beliefs: Links to proficiency. *Modern Language Journal, 79*(3), 372-386.

Oxford, R., & Shearin, J. (1994). Language learning motivation: Expanding the theoretical framework. *Modern Language Journal, 78,* 12-28.

Sarnoff, I. (1970). *Social attitudes and the resolution of motivational conflict.* Harmondsworth: Penguin.

**Appendix 1**

**Dear participants:**

Please read the following items and then select the choice which is more consistent with your perspectives. It is worth mentioning that your choices are just for the purpose of research and the results will be kept confidential.

**Gender:** Male □ Female □

**Age:**

**Educational Background:**

| | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1. Familiarity with the place of the exam helped me perform better. | □ | □ | □ | □ | □ |
| 2. The time of administration affected my performance. | □ | □ | □ | □ | □ |

| | | | | | |
|---|---|---|---|---|---|
| 3. The physical conditions such as light, temperature, chairs affected my performance. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Lack of familiarity with computer use affected my performance negatively. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. Fear of the test affected my performance. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. Familiarity with the test format & rubric helped me perform better. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 7. I prefer iBT to other English proficiency tests. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 8. It is a proficiency test & doesn't evaluate other competencies. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 9. Time was a significant factor regarding my performance. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 10. Listening is the most difficult section of iBT. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 11. Lack of familiarity with British/Australian accents affected my performance adversely. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 12. They speak very fast in the speaking sections of the test. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 13. The quality of voice affected my performance. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 14. Wearing headphones helped me to perform better. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 15. The reason why I missed some of the questions was that I had to answer while listening. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 16. Listening became harder section by section. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 17. Having to answer a variety of questions distracted me. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 18. The Listening section tended to evaluate my knowledge of vocabulary & speed of typing rather than my listening comprehension. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 19. Fear of listening affected my performance adversely. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 20. Listening to English programs made me perform better on listening section. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 21. Command of vocabulary not helped me to perform better. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 22. Reading is the most difficult section of iBT. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 23. The fact that I had to answer reading questions after listening affected my performance. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 24. The variety of questions in the reading section distracted me. | ☐ | ☐ | ☐ | ☐ | ☐ |

| | | | | | |
|---|---|---|---|---|---|
| 25. Lengthy texts helped me locate the answers easily. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 26. Lengthy texts made me exhausted. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 27. One hour is sufficient to answer all the questions. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 28. Texts became more difficult toward the end of the section | ☐ | ☐ | ☐ | ☐ | ☐ |
| 29. Good command of vocabulary items helped me perform better. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 30. Good command of grammar structures helped me perform better. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 31. The vocabulary list option helped me perform better. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 32. Reading a lot of texts before exam helped me perform this section better. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 33. Writing is the most difficult part of the exam. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 34. Lack of familiarity with typing affected my performance. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 35. I was tired because of performance on two other sections. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 36. Lengthy topics distracted me. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 37. Knowledge of vocabulary& grammar helped me in this section. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 38. Not having enough information on the given topic was my main problem. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 39. Having practiced a lot before the exam helped me perform better & faster. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 40. Knowledge of essay writing helped me write better. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 41. The integrated & independent parts were different in terms of difficulty. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 42. Speaking is the most difficult section of the iBT. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 43. The interview section tests one's listening ability rather than speaking one. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 44. Stress was my main problem in this section. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 45. I had self-confidence at the beginning of the section. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 46. Allocated time was a fair criterion based on which my speaking ability be assessed. | ☐ | ☐ | ☐ | ☐ | ☐ |

| | | | | | |
|---|---|---|---|---|---|
| 47. Living in Iran & have a little opportunity to use the language affected my performance. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 48. The integrated & independent tasks were different in difficulty level. | ☐ | ☐ | ☐ | ☐ | ☐ |

**Appendix 2**
**The Interview Items**

| |
|---|
| What is your overall attitude towards iBT? |
| What is your attitude toward the listening section? |
| What is your attitude toward the reading section? |
| What is your attitude toward the writing section? |
| What is your attitude toward the speaking section? |
| What factors affect your performance adversely? |
| What are the advantages of iBT over the paper-based and computer-based versions of TOEFL test? |
| If you want to add anything to your remarks please mention it. |

# A comparative study of composing processes in reading- and graph-based writing tasks

**HUI-CHUN YANG**
*Department of English Instruction, National Hsinchu University of Education, Hsinchu, Taiwan*

**Bio Data:**
Hui-Chun Yang has a Ph.D. from The University of Texas at Austin and has been an instructor of English in the United States and Taiwan. She also had excellent opportunities to teach English to students at the primary, secondary, and post-secondary levels. She is currently an Assistant Professor at National Hsinchu University of Education in Taiwan. Her specializations include language assessment, second language reading and writing, and learner training.

**Abstract**
The study compares EFL writers' processes in composing reading-based writing (RW) and graph-based writing (GW) tasks developed for a university English proficiency exam. Think-aloud protocols and interviews of ten university-level nonnative English-speaking writers were collected to explore writers' composing processes. The results revealed that both types of the tasks require global comprehension of source texts as well as integrative manipulation of available information for writing. Some differences, however, existed across tasks and writers of varying score levels, with the RW tasks eliciting a more interactive and facilitative process than the GW tasks for the higher scoring writers. These results suggested that these tasks might measure different aspects of academic writing ability. Several considerations of the task constructs should apply in properly determining their use in a language test. The findings could be used to provide insights into the nature of RW and GW tasks and contribute to the validity of source-based writing tasks.

*Keywords:* writing assessments, integrated-tasks, reading-based writing, graph-based writing

## Introduction

The ability to integrate sources into writing has been considered important for academic success (Campbell, 1990; Leki & Carson, 1997). Therefore, a plethora of university assignments have involved writing from multiple sources (Horowitz, 1986; Kirkland & Saunders, 1991). In the same vein, writing tasks requiring writers to compose from language input (e.g., reading passages, lectures) or visual input (e.g., graphs, charts, diagrams) have also been increasingly incorporated into the

assessment batteries of a number of language tests (e.g., Test of English as a Foreign Language – TOEFL, Canadian Academic English Language Assessment – CAEL, International English Language Testing System – IELTS, General English Proficiency Test – GEPT) as a means to increase test fairness (Feak & Dobson, 1996; Read, 1990) and foster positive washback effects on learning and teaching (Cumming, Grant, Mulcahy-Ernt, & Powers, 2004; Fox, 2004). Despite their widespread acceptance, criticisms have been leveled against the use of such tasks due to the fact that they may introduce a source of construct irrelevant variance into the assessment (e.g., Charge & Taylor, 1997; Xi, 2005). In this case, writers' reading comprehension and graph comprehension ability may play roles in affecting writing performance, making score use and interpretation difficult.

The present study explores a proposed inclusion of a RW task or GW task other than a writing-only task in an English proficiency exam administered to liberal arts and science majors enrolling in the course of Freshman English. The exam aims to determine students' preliminary English abilities and if further English for Academic Purposes (EAP) support is needed. This change is to respond to a rising call for more authentic writing tests that simulate real-life writing situations. In considering whether RW tasks or GW tasks should be included in the exam, it is necessary to explore the underlying constructs of these tasks. This study sought to investigate the similarities and differences of writers' processes elicited by the RW and GW tasks.

## Review of relevant literature

### Process in Language Testing

Since the late 1980s, language testers have started to consider test-takers' processes to clarify relationships between test tasks and the target language use context. According to Bachman (1990, 2002), investigations of processes involved may provide insight into the construct validity of a language test. Such information helps determine whether test-takers go through the processes expected by test designers or, in other words, if the test actually measures what it proposes to measure. Most process and strategy research has concentrated on test-takers' mental operations in response to selected-response items (e.g., multiple choice, drag-and-drop, cloze) in reading and listening comprehension tests (Anderson, Bachman, Perkins, & Cohen, 1991; Cohen & Upton, 2007; Douglas & Hegelheimer, 2006). Relatively little is known about how test-takers approach constructed-response items (e.g., writing) (Cohen, 1994). To gain a better understanding about the validity of source-based writing tasks, it is necessary to examine how writers interact with the source texts and how their processes may vary across tasks and writers.

### Process on Source-based Writing

A number of previous studies have examined the processes proficient and less proficient writers use in L2 reading-based writing tasks by drawing on constructivist models of discourse synthesis (Spivey, 1984, 1990, 1997; Spivey & King, 1989) in which *organizing, selecting,* and *connecting* processes are identified. In *organizing*, readers/writers refer to the text organization to identify the overall ideas of the text,

and create a text of their own by rearranging chunks of information in the source text. *Selecting* is when readers/writers determine the most important chunks of information from a pool of content units, and subsequently incorporate the selected ones for their writing. During *connecting*, writers make connection between their background knowledge and content in the sources. Plakans (2009), for example, compared L2 writers' processes in two reading-based writing test tasks based on think-aloud protocols. The results showed that proficient writers used significantly more discourse synthesis processes, *connecting* and *organizing* in particular, than their less proficient counterparts. Using a similar approach, Asención (2004) compared processes used by native English speakers, advanced ESL learners, and EFL learners in completing summary tasks. She found that *organizing*, *selecting* and *connecting* occurred less frequently compared to monitoring and planning across groups. In addition to these operations of textual transformation, research on L2 reading-based writing has indicated three other key composing processes including goal-setting, revising, and monitoring (Esmaeili, 2002; Stein, 1990). Overall, these studies contribute an understanding of major reading and writing processes involved in reading-based writing and the connections between writers' L2 proficiency and their corresponding linguistic or cognitive processes.

Another type of source-based writing, GW tasks that include visual sources in the prompt (Bridges, 2010), has rarely been addressed in language testing literature. Fortunately, cognitive psychology theories offer frameworks of graph comprehension and interpretation that may shed light on the interactions between graphical inputs and test-takers. Previous research has identified three central processes that occur successively during graph comprehension: encoding a visually identifiable feature of a graph (e.g., a line sloping upward), interpreting that feature in relation to their knowledge about graphs (e.g., a rising line implies a mounting relationship), and associating specific graph referents to the graph feature (e.g., "human population is increasing") (Bertin, 1983; Carpenter & Shah, 1998; Kosslyn, 1989; Lohse, 1993; Pinker, 1990). Some studies analyzed the perceptual processes people use to make interpretations about a specific graphical format. Carswell, Emery, Lonon (1993) examined participants' processes in responding to a series of line graphs and found that participants constantly engaged in *global productivity* (i.e., an overall trend in a graph) and *local productivity* (i.e., an *x*- or *y*-axis reference, an interpretation for a specific part of a graph). In experiments where participants compared two wedges without values in a pie chart, Gillan and Lewis (1994) found that many participants mentally lifted one wedge over the other to make appropriate comparisons.

Despite the abundance of literature on graphical information processing, only a few studies have addressed graph interpretation in a testing context. Moreover, most of these studies have been conducted in the contexts of listening (Ginther, 2002; Gruba, 1997) and speaking assessments (Katz, Xi, Kim, & Cheng, 2004; Xi, 2005, 2010). Very little research, apart from internal IELTS validation reports (Bridges, 2010; Mickan, Slater, & Gibson, 2000), has addressed these processes involved in writing assessments. Using think-aloud protocols, Mickan, Slater, and Gibson (2000) investigated nine IELTS candidates' test-taking processes and identified three key processes that occurred successively during graph-based writing: *planning prior to*

*writing*, *formulating text*, and *editing*. Questionnaire results reported in Bridges (2010) revealed six processes commonly used to complete IELTS Academic Writing Task 1: macro-planning, organizing, micro-planning, translating, monitoring, and revising. The findings also showed that macro-planning (e.g., goal-setting, task examination) and monitoring occurred more frequently for skilled writers than less skilled writers. The questionnaire used in the study focused mainly on metacognitive self-regulation operations while interactions between graph reading and writing were rarely discussed.

The studies reviewed above show that composing patterns or process types, in general terms, are similar in reading-based writing and graph-based writing. Moreover, L2 proficiency is found to be an important variable that affects process. However, comparisons between processes involved in completing these tasks have not been investigated in previous work. Based on the need of the university English proficiency exam development and the review of prior research, the following research question was proposed: What similarities and differences exist between L2 writers' processes in composing reading-based and graph-based writing tasks?

## Method

### Participants

The participants were recruited on voluntary basis and were introduced to the study by reading a brochure describing the purpose and procedure of the study. The participants were ten full-time undergraduate students enrolled in the course of Freshman English. They have been studying English for at least ten years and had a wide range of English proficiencies based on their writing scores on the RW and GW tasks. Table 1 presents the participant profile.

Table 1
*Participant Profile*

| Writer* | Gender | Age | Discipline | RW score | GW score |
|---|---|---|---|---|---|
| Yifen | F | 19 | Taiwanese | 4 | 4 |
| Jen | F | 21 | Public Health | 3 | 4 |
| Sam | F | 19 | Chemistry | 4 | 4 |
| Dayi | M | 19 | Japanese | 2 | 3 |
| Lee | M | 19 | Chemistry | 3 | 3.5 |
| Feng | M | 20 | Physical Therapy | 2 | 3 |
| Genna | F | 20 | Health Care Management | 2 | 3 |
| Jing | F | 21 | Public Health | 2.5 | 4 |
| Peiling | F | 20 | Health Care Management | 2 | 2 |
| Wei | F | 19 | Japanese | 2 | 3 |

*All names are pseudonyms.

**Tasks**

The tasks that stimulate the academic writing skills of summarizing and synthesizing source materials were designed for the study and for the potential use in the English proficiency exam. Four argumentative writing tasks, one RW and one GW task for two environmental topics (i.e., global warming and ecotourism), were developed. The argumentative genre was selected because it was fairly common in most academic settings (Cumming et al., 2005). Following the suggestion of using more than one source text for source-based writing tasks (Lewkowicz, 1994), two short passages that present opposing viewpoints were included in each RW task. These passages were modified to be similar in text length, organization, and readability based on several criteria: specific main ideas and supporting details for an argument; Flesch Kincaid Grade level between 11 to 12; Flesch Reading Ease score between 40 to 60; and word count between 210 to 250.

Similarly, the GW tasks were created parallel to each other based on three rough standards: obvious trend changes, number of data points, and chart organization. Line graphs were found best in displaying *x*-*y* trends (Carswell, Emery, & Lonon, 1993; Shan, Mayer, & Hegarty, 1999) and supporting global-integration processing (Carswell, 1990; Hollands & Spence, 1992). On the other hand, pie graphs are good for depicting relative proportions of the data (Simkin & Hastie, 1987; Wilkinson, 1999). Line and pie graphs were thus selected in the development of source graphs in attempts to elicit global comparisons of two graphs rather than mere descriptions of *x*- and *y*-axis and data point values.

These tasks were reviewed by three EFL writing specialists and piloted on five potential participants who were also undergraduate freshman students in the university. Several issues including difficult vocabularies, idioms and slangs, sentences and graph structures, and task instructions were addressed before their actual use in the study.

**Data Collection Procedures**

Concurrent think-aloud verbal protocols, pre- and post-interview responses, and written products were collected to address the research questions. Think-aloud approach was used to gain access to the mental log of individuals performing an assigned task (Cohen, 2000; Ericsson & Simon, 1993; Green, 1998; Mickan, Slater, & Gibson, 2000). According to Krapels (1990), the technique provides valuable insights into L2 writers' cognitive operations during writing. Green (1998) also suggests that verbal protocols are a more direct means of "gathering evidence that supports judgments regarding validity than some of the other quantitative methods" (p. 3). By asking the writers to verbalize their thoughts when responding to the tasks, it was possible to investigate the cognitive processes by which writers transform ideas from sources into their own writing. Field notes detailing instances of writers' behaviors (e.g., underlining) and non-verbal expressions (e.g., frowning) were created to facilitate the interpretation of verbal reports. In addition, individual pre- and post-interviews were conducted to provide data for triangulation.

All data were collected in three sessions. The first session began with a brief orientation to the source-based tasks with sample tasks and an interview about writers' backgrounds, reading, graph reading, and writing experience. Then each

writer received instructions on thinking aloud. They watched a video demonstration of the think-aloud approach, and practiced responding to one sample RW and GW task different from the actual tasks. During the practice, writers were reminded to 1) verbalize their immediate thoughts instead of interpreting them, and 2) talk continuously if they fell silent for more than 20 seconds (Ericsson & Simon, 1993; Green, 1998; Olson, Duffy, & Mack, 1985). Also, given that some participants might be concerned about which language to use in the think-aloud session (Manchón, Murphy, & Roca de Larios, 2005), they were instructed to use any language, at any time, with which they felt comfortable. At the same time, feedback was provided to the writers until they became familiar with the technique (Cohen, 2000).

As soon as the writers were ready, they proceeded to the second session. During this session, each participant completed one RW task and one GW task. The study design was counterbalanced on topic order and task order to reduce any possible order effects (see Table 2). Although there was no time limit for completing the tasks, all participants managed to finish them within one hour.

The last interview session took place within two days after the writing session. This semistructured interview explored writers' approaches to complete the tasks and personal reactions toward the tasks. To thoroughly identify the processes involved and consider all sources contributing to the processes, unusual or unexpected comments and pauses were brought up for discussion.

Table 2
*The Study Design*

| Task types | Topic: ecotourism | Topic: global warming |
|---|---|---|
| RW | 1st: Yifen, Jen | 2nd: Sam, Dayi, Lee |
| GW | 1st: Sam, Dayi, Lee | 2nd: Yifen, Jen |
| RW | 2nd: Jing, Peiling, Wei | 1st: Feng, Genna |
| GW | 2nd: Feng, Genna | 1st: Jing, Peiling, Wei |

**Data Transcription and Analysis**

The analyses proceeded in four stages. The first stage involved analyzing the ten writers' composing processes and their interview responses. The audio-recorded verbal protocols were transcribed verbatim, and the transcripts were segmented into "idea units," described by Kroll (1977) as "a chunk of information which is viewed by the speaker/writer cohesively as it is given a surface form … related … to psychological reality for the encoder" (p. 85), for further analysis. The interview data were transcribed and analyzed thematically across interviews. Themes and patterns were identified and categorized to cross-reference with the verbal protocols.

Then two EFL writing specialists (including the researcher) with two to five years of teaching experience explored patterns and identified categories of processes in the data using line-by-line coding approach (Glaser, 1978). In the course of establishing the coding system (Table 3), two objectives were considered: 1) the coding system had to account for twenty sets of protocols as well as to allow comparisons between processes involved in completing the RW and GW tasks, and 2) process theories reviewed earlier were considered in creating categories. Once the categories had been defined, two coders coded a total of 20 protocols independently. In addition, to allow for identification of the composing sequence, the protocols were reexamined for the groupings of the processes and marked according to the composing phases.

Table 3
*Coding System*

| Process | Description |
|---|---|
| Goal-setting | This involves checking, understanding, and interpreting task prompts and instructions. |
| Global planning | This involves identifying major ideas or trends in source materials based on writers' background knowledge about article and graph structures. |
| Local planning | For RW tasks, this relates to understanding source passages by breaking lexical items, phrases, or sentences into parts. For GW tasks, this involves reading data values or describing *x*- and *y*-axes. |
| Selecting | This refers to episodes in which L2 writers go through source materials and selectively draw relevant information from sources to support their writing. |
| Connecting | This is a process in which L2 writers search for a relationship in source materials. |
| Translating | This involves the transformation of thoughts and ideas into writing. |
| Revising | This is a process in which L2 writers adjust their written texts at word, sentence, or essay level as a result of monitoring. |
| Monitoring | This involves reflecting and checking on overall task progress and fulfillment and identifying mechanical issues such as spelling, punctuation, word choice, and syntax. |

While the study was primarily qualitative in nature, quantifying the process data helps provide more information on the trends of the processes within and across tasks and writers. In the third stage, descriptive analyses including frequency counts, percentages, and central tendency statistics were performed to examine the possible differences and similarities which exist in L2 writers' processes when composing the RW and GW tasks.

In the fourth stage, the score pattern of each writer's essay was analyzed to examine the relationship between score and process. The same two experts thus rated the written texts ($n$ = 20) separately. Prior to the actual scoring, a training session was held to provide scoring guidelines including the RW and GW scoring rubrics (see Appendix A) and anchor essays for each score level. Raters first reviewed the scoring rubrics for RW and GW tasks and clarified the traits for rating. Then they practiced rating five essay responses for the RW and GW tasks

respectively and discussed the scoring criteria and discrepancies between each level. As soon as raters reached an agreement on anchor essays of different score levels, they continued to score the entire set of essays. The texts were rated holistically on a scale of 0 to 5. Content, organization, language use were considered in rating the essays. The inter-rater reliability for the raters was found to be substantial for the RW tasks (Kappa = .83, $p < .001$) and GW tasks (Kappa = .83, $p < .003$). Because no two scores assigned by the raters differed by more than one score point, writers' overall writing performance was represented by an average of the two assigned scores.

## Results and Discussion

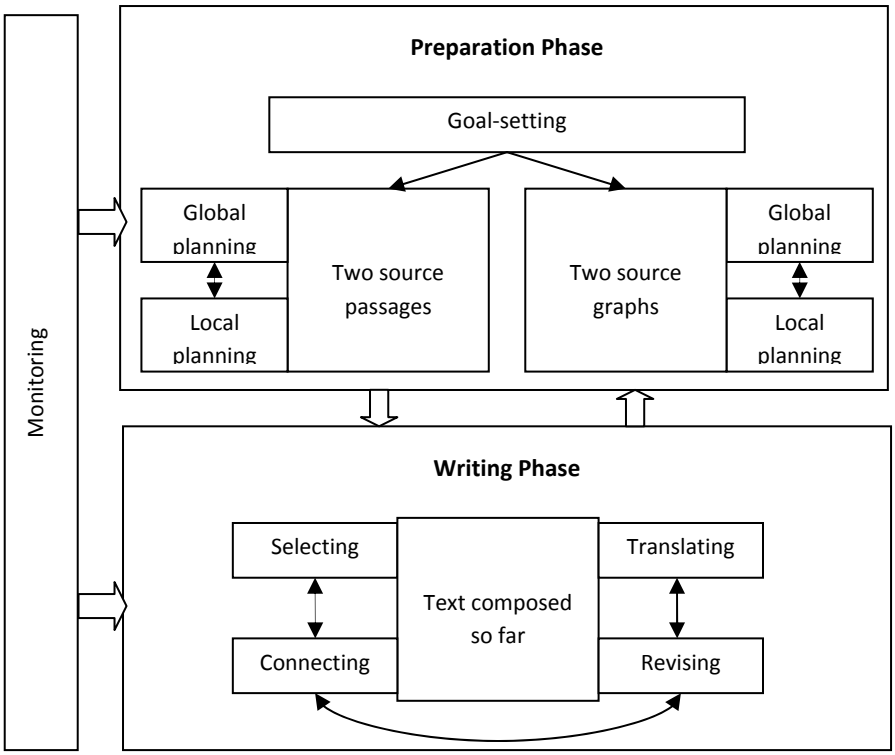
### An L2 Source-based Writing Model

Based on the think-aloud protocols and interview data, a working model (Figure 1) of the writers' composing processes was developed. As Weigle (2002) suggested, L2 writing models should address the contexts in which writer-text interactions take place. This model is a context-specific schematic representation of L2 writers' processes involved in the RW and GW tasks. Although some individual differences were observed, two phases generally occurred for all writers: a preparation phase and a writing phase.

In the preparation phase, all writers read prompts, instructions, and then the source materials. Most of them read prompts and instructions at least once, and then reread some key words and phrases to understand the task goals and purposes further. After examining the task directions, they proceeded to read source passages and graphs using both *global planning* and *local planning*. In the writing phase, the writers reread a portion of source passages or graphs to select information for writing and for comparison between two source materials. When the pieces of information for writing were gathered, they composed their writing and evaluated what had been written. During the preparation phase, the writers read the source texts for the purpose of comprehension, while during the writing phase they reread the source texts for the purpose of summarization. Throughout the two phases, monitoring plays an important role in regulating the use of all processes.

In examining individual writer's process, some writers were found to engage in a more dynamic and constructive process than others. These writers (Sam, Lee, Yifen and Jen) appeared to be those who reported being interested in writing and having daily or weekly blog writing habits. For example, Sam said "It's just like what I often do for blogging. I wrote about what I read and if I am not sure about something, I went back and check for details … I don't write well [sign] but I try very hard." Furthermore, Dayi and Genna who approached the tasks in a more linear fashion described having little experience with writing in their interviews.

*Figure 1.* Composing Processes for RW and GW Tasks




**Composing Processes in the Source-based Writing Tasks**

Table 4 presents the average percentage of each process during the preparation and writing phase across the RW and GW tasks. The subtotal segments coded as process for the two phases revealed that the writers devoted more time and efforts to the preparation phase than the writing phase during the RW tasks. This appeared to be the opposite for the writers responding to the GW tasks. It may suggest that textual sources, rather than graphic sources, place higher demands on writers' cognitive resources for comprehension. The GW tasks, on the other hand, require more cognitive operations during writing than preparation. The following section compares the composing processes involved in the RW and GW tasks.

Table 4
*Average percentage of each process for each type of task*

| | RW | | | GW | | |
|---|---|---|---|---|---|---|
| | *N* | Mean (*sd*) | Range | *N* | Mean (*sd*) | Range |
| ***Preparation phase*** | | | | | | |
| Goal-setting | 60 | .07 (.05) | 12.7 | 59 | .09 (.06) | 17.2 |
| Global planning | 54 | .05 (.04) | 12.8 | 83 | .12 (.05) | 15.1 |
| Local planning | 354 | .36 (.14) | 50.5 | 91 | .15 (.10) | 31.1 |
| *Subtotal* | 468 | .48 | - | 233 | .36 | - |
| ***Writing phase*** | | | | | | |
| Selecting | 174 | .14 (.09) | 25.7 | 103 | .13 (.06) | 22.5 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Connecting | 33 | .03 (.03) | 7.7 | 31 | .03 (.03) | 8 |
| Translating | 220 | .22 (.07) | 18.7 | 169 | .25 (.09) | 31.3 |
| Revising | 21 | .02 (.02) | 4.9 | 56 | .05 (.06) | 15.9 |
| *Subtotal* | 448 | .41 | - | 359 | .46 | - |
| ***Both phases*** | | | | | | |
| Monitoring | 135 | .11 (.09) | 31.9 | 159 | .17 (.11) | 35.5 |

*Note. N* = the total number of think-aloud segment

*a. Preparation phase:* The preparation phase revealed some major differences between the RW and GW tasks. Such differences appeared in the average percentages of *local planning* and *global planning*. *Local planning* occurred at a high rate (36%) for the RW tasks, with participants trying to understand the meaning of source passages by focusing on word-level or phrase-level information. Some writers were found to continuously translating English words into Chinese to capture the main ideas of the passages. *Local planning*, however, occurred at a much lower rate (15%) during the GW tasks. On the contrary, *global planning* occurred at approximately double the rate during the GW tasks (12%) than the RW tasks (5%). As the interviews revealed, the readings provided writers with background ideas and organization for writing so little *global planning* was needed for the RW tasks. However, the GW tasks engaged writers in a greater level of rhetorical and structural planning for writing, suggesting that such tasks would be better in terms of making inferences about a writer's ability to apply logical structures to the content. The following are examples of processes *goal-setting*, *global processing,* and *local processing*:

> *1) '[reads the instructions] So summarize the ideas … I am supposed to write a summary on two essays, okay, about one hundred fifty words, and here are two essays below. So summarize means … to get main ideas. Main ideas from these two passages … passage one and two. Ok.' (Yifen, RW)*
>
> *2) '[reads the graphs] Zhe nian pai fang sheng gao (the emissions have gone up since this year) … chi xu sheng gao dao liang qian nian (and keep going up to two thousand). Zhe bian zui gao (here is the highest) liang qian nian zhi hou kai shi xia jiang (and they went down after two thousand).' (Peiling, GW)*
>
> *3) '[reads a sentence and translates some phrases into Chinese] From heat waves re lang (heat waves) to storms to floods shui zai (floods) to fires to massive glacial glacial … melts, the global climate … di qiu wen du ma (atmospheric temperature)?' (Dayi, RW)*

*b. Writing phase:* During the writing phase, the top two processes that occurred most frequently were *translating* and *selecting*. *Translating* occurred at a highest rate in this phase for both RW (22%) and GW (25%) tasks. *Selecting* was the next most frequently observed process for the RW (14%) and GW (13%) tasks. The number of think-aloud segment revealed that writers were engaged in more *selecting* process when composing the RW tasks than the GW tasks. This may suggest a more

dynamic knowledge selecting and transforming process. Several writers described in their interviews that they carefully chose some pieces of information that represented the essence of the passages and deliberately neglected the others. The result indicates that the RW tasks can better elicit writers' ability to assess the values of information and make informed decisions on what to cite in their writing. Examples of processes *selecting, linking, translating,* and *revising* include:

> *4) '[rereads source graphs] Okay high point I need to write about high point. Let me look … here … a high point … very high here in year … two thousand the number of tourists is about … about thirty thousand.' (Lee, GW)*
>
> *5) '[rereads source passages] The first passage talks about ecotourism … the second one also about ecotourism … the same thing … but they are different, different in some way. They have different points … one agrees this is a good idea … but the other don't think so.' (Jing, RW)*
>
> *6) '[writes down a sentence] Many scientists' prediction is not very well. Prediction for the problem is not right.' (Dayi, RW)*
>
> *7) 'The number of tourists are more than twenty thousand … no shi tai (tense) cuo le (wrong) [crosses out a word] … ying gai shi (should be) 'was' twenty thousand.' (Feng, GW)*

*c. Monitoring in preparation and writing phase:* The *monitoring* process occurred at a higher rate during the GW tasks (17%) than the RW tasks (11%). For example:

> *8) '[rereads source graphs] I don't understand this graph [frowns]… why? [rereads instructions] Make connections? This graph and these two here. It says it's about CO2 emissions between these two years. How about here? One is nineteen-ninety and this one two thousand. Oh so here these two highs and these two years. Got it! I am so smart.' (Jing, GW)*

**Variation in Composing Processes by Performance Level**

To explore variation in composing processes among writers, RW and GW essays were scored to compare processes used at different performance levels. The writers' scores ranged from 2 to 4 with all half-point scores rounded up. For the RW task, five writers were at level 2, three at level 3, and two at level 4. For the GW task, one writer was at level 2, four were at level 3, and five were at level 4. The descriptive statistics for ratings of RW and GW task performance (Table 5) show that writers' mean scores on the GW tasks, regardless of the topics, were generally higher than those on the RW tasks, suggesting that the RW tasks seem to be more challenging than the GW tasks.

Table 5
*Descriptive Statistics for the Ratings of RW and GW Task Performance*

| Task types and topics | Mean scores | Minimum | Maximum | Standard deviation |
|---|---|---|---|---|
| RW | | | | |
| Ecotourism | 2.7 | 2 | 4 | .84 |
| Global warming | 2.6 | 2 | 4 | .89 |
| GW | | | | |
| Ecotourism | 3.3 | 3 | 4 | .44 |
| Global warming | 3.4 | 2 | 4 | .89 |

When considering the range of each writer's composing processes, no consistent pattern is presented for essays scored as 4 and the other two levels; however, clear differences were found in the process means across score levels (Table 6). A comparison of the process means across three levels shows that writers at the highest level were engaged in approximately one third more and one fifth more processes than writers at the lower two levels for the RW and GW task respectively. Another difference lies in the score distribution. For the RW task, the scores cluster at a lower end whereas the opposite is true for the GW task. A further examination of individual writer's interview data reveal that three writers, Jen, Lee, and Jing who scored 3 in the RW tasks but 4 in the GW tasks had much experience with graphs. They stated that lots of their academic coursework involved analyzing graphs and converting data into graphs, which might indicate that writers' familiarity with graphs can play an important role in their performance on the GW tasks.

Table 6
*Range and Means of Processes Used by Each Score Level (n = 10)*

| | RW Process | | | | | RW Mean | GW Process | | | | | GW Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| L 4 | 139 | 148 | | | | 143.5 | 64 | 30 | 186 | 55 | 87 | 84.4 |
| L 3 | 81 | 141 | 78 | | | 100 | 53 | 113 | 36 | 69 | | 67.75 |
| L 2 | 104 | 169 | 80 | 41 | 70 | 92.8 | 62 | | | | | 62 |

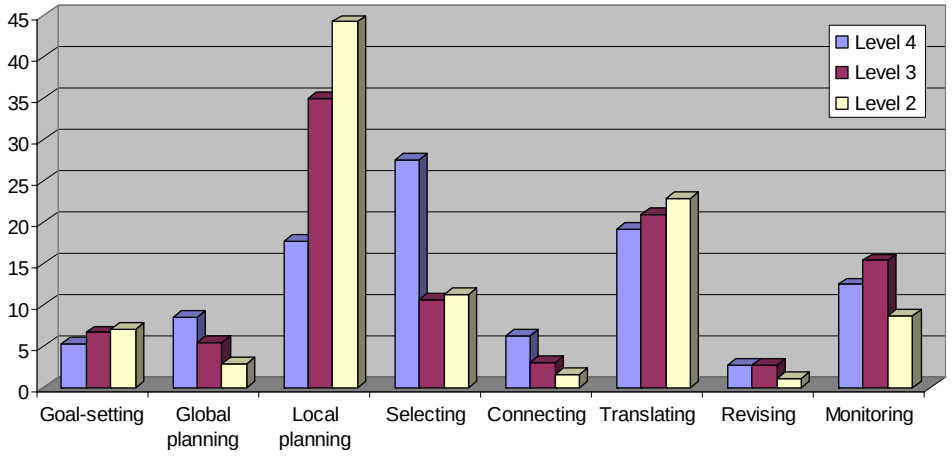
*Note:* L = score level. The numbers under the RW Processes and GW Processes columns refer to the total number of processes used by different score levels (e.g., writers who obtained four points for their RW tasks engaged in 139 and 148 composing processes during writing). RW mean and GW mean refer to the average of RW and GW processes used by each score level.

Previous research has shown that what distinguishes writers are the *range* as well as the *type* of processes involved in writing (Asención, 2004; Cohen, 1994). Figure 2 presents the occurrence percentage of each type of process engaged during the RW tasks. Several key patterns were found. The highest scoring writers engaged

in more global processes, such as identifying text structures and skimming for key ideas, which confirms previous reading-writing research (Cohen, 1994; Plakans, 2009). These writers also purposefully selected major ideas for writing and connected one piece of information with another from different source texts. The prevalent use of *global planning*, *selecting*, and *connecting* among higher scorers may suggest these were facilitative processes for the RW tasks.

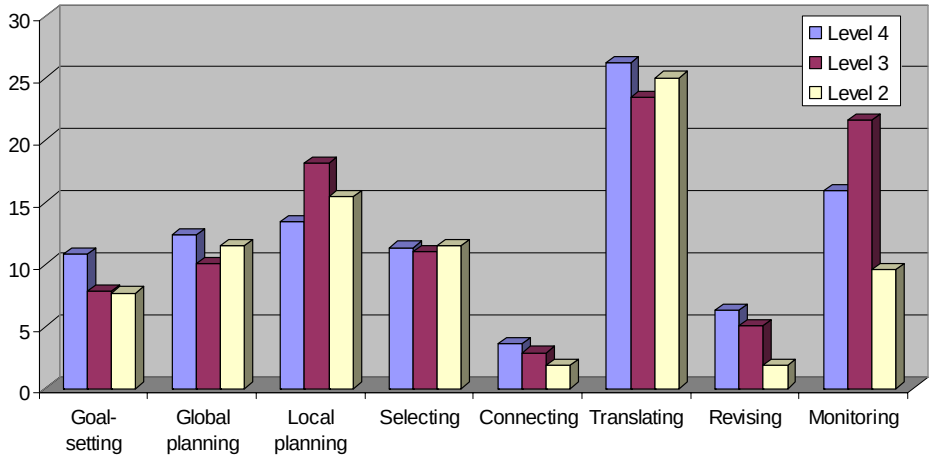
On the other hand, the writers at level 3 and 2 focused more on local-level information. They broke words and sentences into parts and resorted to translation frequently as a means to understand the whole texts, which might suggest that they found the source passages challenging than the higher scoring writers. Genna, Peiling, and Wei, who scored 2 in the RW tasks, had mentioned in their interviews of being 'uncertain,' 'confused,' and 'frustrated' about the source passages.

*Figure 2.* Percentages of Different Types of Composing Processes Used by Each Score Levels (RW)




Similar patterns were found for the GW tasks (Figure 3) while the differences between levels were fairly small compared to those in the RW tasks. Take *local planning* for example. Writers at level 4 in the RW tasks used this process much less frequently than the rest of the writers, but the difference was minor in the case of the GW tasks. This appeared to reflect the nature of these tasks. The RW tasks require a mastery of basic word- or sentence-level comprehension while the GW tasks do not. Yet what is not revealed in the figure is the language difficulties common to writers at level 2 and 3 in completing the GW tasks. Dayi, Genna, and Wei had indicated in their interviews that they felt the GW tasks were more challenging because they had to transform numerical data and visual trends into written texts. They often found difficulties in searching for appropriate trend-describing words (e.g., rise, fall, fluctuate) for writing. In contrast, Jen, Sam, Lee, and Feng, science majors who scored 4 in the GW tasks stated in their interviews that lots of their academic coursework involved analyzing graphs and converting data into graphs. These writers found their experience had helped them determine what to write and how to organize the content.

*Figure 3.* Percentages of Different Types of Composing Processes Used by Each Score Levels (GW)




## Conclusion

The goal of the study was to investigate L2 writers' composing processes during source-based tasks to explore their underlying constructs. The two sets of tasks were designed and used in the study to measure academic writing skills. Concurrent verbal protocols were gathered to examine how writers created meaning through the transformation of source texts. The findings may enhance our understanding of the possible constructs of academic writing, and provide test designers and users more information for the development and interpretation of these test tasks.

The source-based writing model suggests that the writers approach both tasks with the types of problem-solving and discourse-synthesizing processes described in previous literature of source-based writing (Bridges, 2010; Hirvela, 2004; Spivey, 1997), which provides some evidence for construct validity (Messick, 1989) of these tasks. In addition, contrary to the purely linear (Rohman, 1965) and recursive view of writing (Flower & Hayes, 1981; Witte, 1985), the data reveal both linear and iterative nature of composing from sources depending on the writers' *procedural knowledge*, a repertoire to manage a range of thinking operations for the purpose of achieving the writing goals, as well as their *task environment*, the distance between writers' text composed so far and expectations they hold for themselves (van den Bergh & Rijlaarsdam, 2001).

The results on the proportions of processes engaged during the RW and GW tasks also reveal the nature of these tasks. The RW tasks minimally require a mastery of threshold reading comprehension skill at the sentence and paragraph levels. They are better in capturing writers' ability to evaluate sources and make selections for use in writing. Such ability is particularly important since authors almost always draw on ever-growing pools of information in academic writing. On the other hand, the GW tasks require writers to sequence the content in a unified logical structure as well as to compose using trend-describing vocabulary, which might pose some difficulties for writers who are not familiar with graphs.

When comparing the processes engaged by writers at different score levels, the data indicate that the highest scoring group showed a tendency to use facilitative

processes more frequently than the mid- and low-scoring groups during the RW tasks, a finding that can be linked to research on cognitive operations during reading-based writing (Connor & Kramer, 1995; Yang & Shi, 2003). Such differences, however, were not evident for the GW tasks. It seems that the RW tasks prompted more constructive processes than the GW tasks for the more proficient writers. These findings also reveal that these tasks may have been measuring different aspects of source-based writing ability. This assumption is further verified by the results that not all of the writers scoring high in the RW tasks scored high in the GW tasks. The three writers who scored high in the GW tasks but not in the RW tasks happened to be those who had much experience with graph interpretation and analysis. Clearly, graph familiarity and comprehension ability play roles in graph-based writing performance. As indicated in a number of studies on graph-based assessment tasks (Katz, Xi, Kim, & Cheng, 2004; Xi, 2005), caution must be taken due to the potential influence of graph familiarity on writers' processes and performance.

Given that GW tasks are designed to assess students regardless of their background knowledge on graphs, they were not considered to be included in the current English proficiency exam. However, GW tasks may have potential for use in assessing writing ability of science majors considering that visual literacy increasingly has been viewed as a prerequisite for understanding academic texts and graphicacy skills have started to be seen as part of the larger construct of academic writing (Hyland, 2006; Kress, 2003).

Several limitations of this study should be noted. First, the study only focused on four source-based writing tasks specifically developed for a university English proficiency exam. Task and topic effects may occur considering there are many variants of academic writing. In addition, although the think-aloud method can be effective for understanding writers' composing processes, they may affect or alter writers' thought processes and performances. Also, writers' processes would have been different if they participate in the actual proficiency exam where they would be limited in time. Finally, the study is exploratory in nature and only considers ten participants; thus, future studies incorporating more participants as well as participants with different backgrounds would provide greater insight into the nature of these tasks. In light of the restricted scope of this study, the results should be interpreted with caution.

**References**

Anderson, N. J., Bachman, L. F., Perkins, K., & Cohen, A. D. (1991). An exploratory study into the construct validity of a reading comprehension test: Triangulation of data sources. *Language Testing, 8*(1), 41–66.

Asención, Y. (2004). *Validation of reading-to-write assessment tasks performed by second language learners* (Unpublished doctoral dissertation). Northern Arizona University, Flagstaff, AZ.

Bachman, L. F. (1990). *Fundamental consideration in language testing*. Oxford: Oxford University Press.

Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing, 19*(4), 453–476.

Bertin, J. (1983). *Semiology of graphics: Diagrams networks maps*. Madison, WI: University of Wisconsin Press.

Bridges, G. (2010). Demonstrating cognitive validity of IELTS academic writing task 1. *Research Notes, 42*, 24–33.

Campbell, C. (1990). Writing with others' words: Using background reading text in academic compositions. In B. Kroll (Ed.), *Second language writing* (pp. 211–230). Cambridge, UK: Cambridge University Press.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied, 4*, 75–100.

Carswell, C. M. (1990). *Graphical information processing: The effects of proximity compatibility*. Paper presented at the Human Factors Society 34th Annual Meeting, Santa Monica, CA.

Carswell, C. M., Emery, C., & Lonon, A. M. (1993). Stimulus complexity and information integration in the spontaneous interpretation of line graphs. *Applied Cognitive Psychology, 7*(4), 341–357.

Charge, N., & Taylor, L. B. (1997). Recent developments in IELTS. *ELT J, 51*(4), 374–380.

Cohen, A. D. (1994). English for academic purposes in Brazil: The use of summary tasks. In C. Hill, & K. Parry (Eds.), *From testing to assessment: English as an international language* (pp. 174–204). London, England: Longman.

Cohen, A. D. (2000). Exploring strategies in test taking: Fine-tuning verbal reports from respondents. In G. Ekbatani, & H. Pierson (Eds.), *Learner-directed assessment in ESL* (pp. 131–145). Mahwah, NJ: Erlbaum.

Cohen, A. D., & Upton, T. A. (2007). 'I want to go back to the text': Response strategies on the reading subtest of the new TOEFL. *Language Testing, 24*(2), 209–250.

Connor, U. M., & Kramer, M. G. (1995). Writing from sources: Case studies of graduate students in business management. In D. Belcher, G. Braine, & D. Dewhurst Belcher (Eds.), *Academic writing in a second language* (pp. 155–182). Norwood, NJ: Ablex Publishing Corporation.

Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. (2004). A teacher-verification study of speaking and writing prototype tasks for a new TOEFL. *Language Testing, 21*(2), 159–197.

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Difference in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing, 10*, 5–43.

Douglas, D., & Hegelheimer, V. (2006, June 30). *Learner strategies and knowledge in TOEFL Listening Tasks: A validation study.* Paper presented at the Language Testing Research Colloquium, Melbourne, Australia.

Ericsson, K. A., & Simon, H. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Esmaeili, H. (2002). Integrated reading and writing tasks and ESL students' reading and writing performance in an English language test. *The Canadian Modern Language Review 58*(4), 599–622.

Feak, C., & Dobson, B. (1996). Building on the impromptu: A source-based academic writing assessment. *College ESL, 6*(1), 73–84.

Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication, 32*(4), 365–387.

Fox, J. (2004). Test decisions over time: Tracking validity. *Language Testing, 21*(4), 437–465.

Gillan, D. J., & Lewis, R. (1994). A componential model of human interaction with graphs: 1. Linear regression modeling. *Human Factors, 36*(3), 419–440.

Ginther, A. (2002). Context and content visuals and performance on listening comprehension stimuli. *Language Testing, 19*(2), 133–167.

Glaser, B. G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press.

Green, A. J. F. (1998). *Using verbal protocols in language testing research: A handbook*. Cambridge, UK: Cambridge University Press.

Gruba, P. (1997). The role of video media in listening assessment. *System, 25*, 335–346.

Hirvela, A. (2004). *Connecting reading and writing in second language writing instruction*. Ann Arbor, MI: The University of Michigan Press.

Hollands, J. G., & Spence, I. (1992). Judgements of change and proportion in graphical perception. *Human Factors, 34*(3), 313–334.

Horowitz, D. (1986). What professors actually want: Academic tasks for the ESL classroom. *TESOL Quarterly, 20*(3), 445–462.

Hyland, K. (2006). *English for academic purposes: An advanced resource book*. New York, NY: Routledge.

Katz, I. R., Xi, X., Kim, H., & Cheng, P. C.-H. (2004). *Elicited speech from graph items on the test of spoken English. TOEFL research report no. 74*. Princeton, NJ: Educational Testing Service.

Kirkland, M. R., & Saunders, M. A. P. (1991). Maximizing student performance in summary writing: Managing cognitive load. *TESOL Quarterly, 25*(1), 105–121.

Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology, 3*, 185–225.

Krapels, A. R. (1990). An overview of second language writing process research. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 37–56). Cambridge, UK: Cambridge University Press.

Kress, G. (2003). *Literacy in the new media age*. London: Routledge.

Kroll, B. (1977). Combining ideas in written and spoken English: A look at subordination and coordination. In E. O. Keenan, & T. L. Bennett (Eds.), *Discourse Across Time and Space* (pp. 69–108). Los Angeles, CA: University of Southern California.

Leki, I., & Carson, J. (1997). Completely different worlds: EAP and the writing experiences of ESL students in university courses. *TESOL Quarterly, 31*(1), 39–69.

Lewkowicz, J. (1994). Writing from sources: Does source material help or hinder students' performance? In N. Bird (Ed.), *Language and Learning* (pp. 204–217). Hong Kong: Institute of Language in Education.

Lohse, G. L. (1993). A cognitive model of understanding graphical perception. *Human-Computer Interaction, 8*, 353–388.

Manchón, R. M., Murphy, L., & Roca de Larios, J. (2005). Using concurrent protocols to explore L2 writing processes: Methodological issues in the collection and analysis of data. In P. K. Matsuda, & T. Silva (Eds.), *Second language writing research: Perspectives on the process of knowledge construction* (pp. 191–205). Mahwah, NJ: Erlbaum.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–104). New York: American Council on Education & Macmillan.

Mickan, P., Slater, S., & Gibson, C. (2000). Study of response validity of the IELTS writing subtest. In R. Tulloh (Ed.), *International English Language Testing System* (Vol. 3, pp. 29–48). Canberra, Australia: IELTS Australia.

Olson, G. M., Duffy, S. A., & Mack, R. L. (1985). Question asking as a component of text comprehension. In A. C. Graesser, & J. B. Black (Eds.), *The psychology of questions* (pp. 219–226). Hillsadale, NJ: Erlbaum.

Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73–126). Mahwah, NJ: Erlbaum.

Plakans, L. (2009). Discourse synthesis in integrated second language writing assessment. *Language Testing, 26*(4), 561–587.

Read, J. (1990). Providing relevant content in an EAP writing test. *English for Specific Purposes, 9*, 109–121.

Rohman, D. G. (1965). Pre-writing: The stages of discovery in the writing process. *College Composition and Communication, 16*, 106–112.

Shan, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology, 91*(4), 690–702.

Simkin, D., & Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association, 82*, 454–465.

Spivey, N. (1984). *Discourse synthesis: Constructing texts in reading and writing.* Newark, DE: International Reading Association.

Spivey, N. (1990). Transforming texts: Constructive processes in reading and writing. *Written Communication, 7*, 256–287.

Spivey, N. (1997). *The constructivist metaphor: Reading, writing, and the making of meaning*. New York: Academic Press.

Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly, 24*, 7–26.

Stein, V. (1990). Exploring the cognition of reading-to-write. In L. Flower, V. Stein, J. Ackerman, M. J. Kantz, K. McCormick, & W. C. Peck (Eds.), *Reading-to-write: Exploring a cognitive and social process* (pp. 119–143). New York, NY: Oxford University Press.

van den Bergh, H., & Rijlaarsdam, G. (2001). Changes in cognitive activities during the writing process and relationships with text quailty. *Educational Psychology, 21*, 373–385.

Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.

Wilkinson, L. (1999). Graphs for research in counseling psychology. *The Counseling Psychologist, 27*(3), 384–407.

Witte, S. P. (1985). Revising, composition theory and research design. In S. Witte (Ed.), *The acquisition of written language: Response and revision* (pp. 250–284). Norwood, NJ: Ablex.

Xi, X. (2005). Do visual chunks and planning impact performance on the graph description task in the SPEAK exam? *Language Testing, 22*(4), 463–508.

Xi, X. (2010). Aspects of performance on line graph description tasks: Influenced by graph familiarity and different task features. *Language Testing, 27*(1), 73–100.

Yang, L., & Shi, L. (2003). Exploring six MBA students' summary writing by introspection. *Journal of English for Academic Purposes, 2*, 165–192.

**Appendix A:** Scoring rubrics for the RW and GW tasks

| Score Level | Task Description |
|---|---|
| **5** | A response at level 5<br>• is effective in selecting major information from two source texts/graphs to support one another and connecting relevant ideas<br>• demonstrates unity, coherence, syntactic variety, and appropriate word choice<br>• contains minor lexical or syntactical errors that do not interfere with meaning |
| **4** | A response at level 4<br>• is effective in selecting and connecting major information from two source texts/graphs although some ideas may not be fully elaborated<br>• demonstrates unity, coherence, syntactic variety, and appropriate word choice although it may contain few unclear connections or occasional redundancy<br>• contains few lexical or syntactical errors that do not interfere with meaning |
| **3** | A response at level 3<br>• contains some but not all major points from two source texts/graphs and the points are imprecisely or incorrectly presented or connected<br>• demonstrates unity and coherence although it may contain few obscure connections and imprecise word choice<br>• displays limited syntactic structures and vocabulary<br>• contains some lexical or syntactical errors that occasionally obscure |

| | |
|---|---|
| | meaning |
| **2** | A response at level 2<br>• contains limited relevant points from two source texts/graphs and they are significantly misrepresented<br>• displays little organization or inadequate connections of ideas<br>• contains inappropriate word choice<br>• displays many lexical or syntactical errors that largely obscure meaning |
| **1** | A response at level 1<br>• contains little or no relevant information from two source texts/graphs<br>• is disorganized and underdeveloped<br>• displays serious and frequent lexical and syntactical errors that make understanding of the writing unlikely |
| **0** | A response at level 0<br>• contains copied words from the source passages<br>• is written in a foreign language<br>• is left blank |

# Hybrid Modeling of Intelligence and Linguistic Factors as Predictors of L2 Writing Quality: A SEM Approach

**REZA PISHGHADAM**
*Associate Professor of TEFL, Ferdowsi University of Mashhad*

**MOHAMMAD ALI SHAMS**
*PhD Candidate of TEFL, Ferdowsi University of Mashhad*

**Bio Data:**
Reza Pishghadam is an associate professor in TEFL in Ferdowsi University of Mashhad, Iran. His research interests are psychology of language education and sociology of language.

Mohammad Ali Shams is a PhD Candidate of TEFL in Ferdowsi University of Mashhad, Iran. His research interests are language assessment and the applications of cognitive science in TEFL.

**Abstract**
The aim of this paper was to investigate the role of linguistic and intelligence factors in L2 writing. The sample included 347 Iranian learners of English. Six tests were administered to measure the participants' grammar knowledge, depth of vocabulary knowledge, breadth of vocabulary, verbal intelligence, narrative intelligence, and writing ability. Two SEM models were compared to each other. Model 1 only included grammar knowledge, depth and breadth of vocabulary knowledge as the linguistic factors of writing. Model 2, proposed by the researchers, included verbal and narrative intelligences as well. The models were then linked to the data to see which one fits better. The results of structural equation modeling show that Model 2 has better fit indices producing better parameter estimates. In the end, the applications and implications of the findings for L2 writing pedagogy and assessment are discussed.

*Keywords:* L2 writing, Linguistic factors, Intelligence factors, Structural equation modeling

## Introduction

Writing is frequently labeled as the most difficult skill to master for language learners (e.g. see Berman & Cheng, 2001; Erkan & Saban, 2011; Snider, 2002). Coping with the difficulties observed in teaching writing would not be likely if one does not have an in-depth understanding of the nature and dynamics of this construct (L2 writing ability). Understanding the cognitive processes involved in writing tasks

while learning a second or foreign language seems necessary for tackling the problems observed in writing classrooms. Identifying, describing, and explaining the cognitive factors involved in L2 writing have captured scholars' attention from a range of disciplines such as neurolinguistics (Barnes, Dennis, & Hetherington, 2004; Paradis & Hildebrandt, 1985; Weekes, Yin, Su, & Chen, 2006), second language acquisition (Bialystok, 2002; Sandberg & Hjelmquist, 1996), teaching English as a foreign language (Escribano, 1999; Gupta & Woldermariam, 2011; Nakamaru, 2010), discourse analysis (Hyland, 2008), and narrative psychology (Bloome, Katz, & Champion, 2003). Expectedly, applied linguists can use the findings of cognitive sciences for solving the learners' problems in L2 writing.

Widdowson (2000) warned applied linguists by pointing out the necessity of attending the practical problems of language learners instead of getting lost in theoretical mazes built out of academic jargon and abstract concerns. This view has been accepted by many writing researchers (e.g., Duong, Cuc, & Griffin, 2011; Ferris, 2010; Lei, 2008; Schneider, 2011). The purpose of any theoretical model for explaining the internal mechanisms of writing must be solving the problems of learning and teaching. The major problem with the learning of writing skill in language classrooms is that learners are taught grammar and vocabulary but fail to translate these newly achieved competencies into their writing performance to the expected extent (Olinghouse & Leaird, 2009; Wolsey, 2010; Zhou, 2009). One can ask why syllabi for teaching writing designed based on lexical and syntactic notions do not function as they should.

The authors believe the findings of cognitive sciences shed more light on the practical problems of learning L2 writing; welcoming the ideas and research tools from other disciplines will let writing research flourish and prosper more than ever. An interdisciplinary momentum is needed to accelerate the writing research in the right direction. As one of the distinguished abilities of human intellectual enterprise, writing must be investigated from perspectives which are not merely limited to the theoretical boundaries of linguistics. The idea of the insufficiency of purely linguistic accounts of writing and the need for adopting an interdisciplinary approach to the study of learners' problems with writing first came to the researchers during reflecting on practical problems, and then the review of the related literature let it gradually evolve into a testable hypothesis about the role of intelligence factors in developing writing ability. Although the literature of writing research is almost dominated by the conventional view based on which the writing construct is solely affected by linguistic competencies, signs of interdisciplinary solutions to writing problems can also be tracked down. This will be discussed in the following sections. This study was launched with such perspective.

The central hypothesis of this study is that adding intelligence factors especially narrative intelligence to the traditionally acclaimed linguistic factors will create a more realistic image of writing ability and its internal mechanisms. In fact, the core idea of the present study originated from the intuitive realization of the insufficiency of linguistic competencies for explaining the dynamics of writing ability; this idea was inspired by close observation and careful assessment of synchronic and diachronic changes in language learners' writing performance. Reviewing the related literature (see Goldberg, Schwarz, & Porat, 2011; Gustilo,

2010; Lee & Tan, 2010; Pantaleo, 2010; Randall, 1999) provided a theoretical framework to formulate a plausible hypothesis: a model of factors influencing L2 writing is more explanatory if it includes intelligence factors. Therefore, the main question addressed in the present study reads as follows:

- Does a model of writing with intelligence and linguistic factors fit the learners' writing scores better than a model that only includes linguistic factors?

## Theoretical Framework

This section presents a brief literature on different types of factors influencing one's L2 writing quality. First, the priorities of teaching writing reflected in the work of writing researchers are introduced. Then, the focus on higher-order processes involved in writing is elaborated and justified based on the most recent developments in L2 writing research. In the next step, more articulate accounts of the role of cognitive factors in L2 writing are presented. Finally, adopting an interdisciplinary approach, the role of intelligence factors in developing the writing ability is discussed.

### Linguistic vs. Cognitive Factors in L2 Writing

What are the teaching priorities in a writing classroom? The writing researchers can be arguably divided into two major groups considering their answer to the above-mentioned question. The first group (e.g. Truscott, 1996; Ferris, 1999, 2004; Nakamaru, 2011) considers learners' knowledge of grammar and vocabulary as the two main factors that count and should be attended by the teacher while the second group (e.g. Devine, Railey & Boshoff, 1993; Hamp-Lyons & Mathias, 1994; Skehan & Foster, 2001; Robinson, 2005) believe that there are non-linguistic higher order processes which should not be overlooked by the teachers. A line of debate which is formed within the first group can be traced in the long-lasting controversy over the superiority of lexical or syntactic feedback in teaching writing.

Truscott (1996) took a strong stance indicating that grammar correction in L2 classrooms is harmful and should be abolished. In response to this view, Ferris (1999) argued for the benefits of error correction claiming that Truscott (1996; 1999) has overlooked the positive evidence on the effects of teachers' syntactic feedback to L2 writers. Ferris (2004) follows the same line of reasoning, and while reminding the readers of the positive effects on error correction argues for the insufficiency of the research on syntactic feedback hence the unviability of any conclusive stance on the issue. What matters is that Ferris (2004) considers syntactic feedback as a priority for improving the learners' writing ability. He does refer to the existence of some higher-order processes involved in writing but does not provide any details regarding their dynamics or any possible interactions between these non-linguistic factors and the syntactic component of writing.

Nakamaru (2011) discusses the syntactic and lexical feedback provided by tutors in writing centers. Tutors, in accordance with the policy of these centers, usually focus on higher-order aspects of the written texts allocating less time to the linguistic details. She believes that the current writing methodology is loaded with

too much emphasis on sentence level feedback which overlook important problems in learners' writing ability; she also asserts that when attending the micro features of the texts, the tutors should not spend too much time for syntactic nuances of writing while the learners are eager to strengthen the lexical aspects of their writing. This view favors a lexical syllabus for teaching writing, which sees grammar as a secondary teaching priority. In fact, it sees the written text as a body whose structure is built up by grammatical patterns and is fleshed by learners' vocabulary. In Nakamaru's (2011) opinion, this type of syntactic feedback leads to the production of "vague and confusing" sentences in the students' writings (p. 98). One important point which is usually overlooked in such debates is that the nature of some cognitive factors or as Ferris (2004) put it, higher-order processes in writing, is different from linguistic factors such as grammar and vocabulary knowledge.

**Cognitive Accounts for L2 Writing Tasks**

The role of cognitive factors in improving and also hindering the writing ability has been a frequent theme in writing research during the past two decades. Devine, Railey, and Boshoff (1993) discussed the implications of cognitive models for L1 and L2 writing. They showed that writers' knowledge of personal, task, and strategy variables are highly interactive, and altogether they form one's cognitive model of a cognitive task. Hamp-Lyons and Mathias (1994) found that, in contrast to the common belief, expository and personal prompts were associated with lowest writing scores; the learners received the highest scores in response to argumentative and public prompts. They concluded that higher cognitive task complexity stimulates the students more strongly hence the higher wiring scores.

Kuiken and Vedder (2008) compared Skehan and Foster's (2001) Limited Attentional Capacity Model with Robinson's (2005) Cognition Hypothesis to see which one fits the writing data better. According to Skehan and Foster's Model, when under pressure, the brain prioritizes meaning over form of the language. Therefore, they predict that in more cognitively complex tasks the learners are likely to achieve lower scores. However, according to Robinson's model, increase in task complexity does not degrade linguistic output because cognitive factors are associated with different resource pools and can work parallel to each other. Kuiken and Vedder (2008) found support for the latter model because written products of cognitively more demanding task were found to be more accurate (with lower error ratio per T-unit) while syntactic complexity and lexical variation were not affected by cognitive task complexity. In another attempt to investigate the cognitive dynamics of L2 writing tasks, Ong and Zhang (2010) defined two types of writing fluency and three types of cognitive task complexity in their study of L2 writing. They found that increasing cognitive task complexity with respect to planning time continuum creates more writing fluency type II (mean number of words produced per minute) and lexical complexity.

Although none of the above scholars makes any explicit reference to the cognitive or intelligence factors affecting writing ability, their results point out the significance of one's cognitive abilities or intelligences in the process of writing. It seems that, instead of measuring the cognitive abilities by the use of validated psychometric scales, writing researchers prefer to measure learners' performance on

different aspects of the writing ability in response to tasks with different levels of cognitive complexity (see Kormos, 2011; Kuiken & Vedder, 2007; Wolsey, 2010). Although their evading of straight measurement of intelligence factors may be justified on logistical grounds, one cannot deny that this is achieved by compromising the psychometric solidarity and theoretical independence of writing models. In other words, measuring learners' writing response to cognitively complex tasks cannot replace measuring cognitive abilities which are assumed to be independent of the writing process.

The overwhelming presence of cognitive factors has derived the researchers to give up the linguistic attachments of writing and take it as a cognitive ability which can be realized in both languages. Hirose (2006) came up with a similar result and tentatively concluded that same writer can choose different organizational patterns regardless of the language. From this perspective, grammar and vocabulary knowledge cannot predict writing ability alone simply because this ability goes beyond the borders established by linguistic competences (see Martinez, Kock, & Cass, 2011). It was the same perspective that originated the main hypothesis of the present study based on which linguistic factors namely grammar knowledge, and depth and breadth of vocabulary knowledge are not enough for explaining the variance observed in foreign language learners' writing performance; in other words, intelligence factors should not be excluded from writing models anymore. Intelligence factors do play a role in developing language proficiency particularly writing ability (see Eng & Mustapha, 2010; Rahimi & Qannadzadeh, 2010). The appearance of verbal, emotional, and narrative intelligences in the literature of language learning and writing research during the last decade marks an interdisciplinary trend which seeks new solutions for the long-standing problems of teaching writing.

The study of the relationship between verbal and emotional intelligences and L2 writing ability is a recent trend in applied linguistics. In some studies, the place of writing is limited to a marginal role and its dynamics are not discussed in lengths. For example, Fahim and Pishghadam (2007) studied the role of emotional, psychometric, and verbal intelligences in the academic achievement of university students majoring in English. The academic achievement was measured by the students' scores in several courses particularly English (L2) writing. They found that IQ has little predictive validity for academic success while EQ showed a strong relationship with academic success. They also found that verbal intelligence of university students has a meaningful relationship with their academic success. Pishghadam and Ghonsooly (2008) investigated the role of emotional intelligence in second language learning success and found significant relationships between intelligence factors and linguistic factors but did not claim any causal relationships between those variables. In the two above-mentioned studies, academic success was taken as a general construct and the details of the relationship and specific language skills were not discussed, whereas Abiodun and Folaranmi (2007) aimed to investigate such relationship and found that verbal ability has a significant effect on second language writers' achievement in essay writing. Yet Pishghadam (2009) reinforced this line of research by finding causal relationships between verbal and emotional intelligences and the number of errors and writing ability of language

learners. According to the results of his study, Pishghadam (2009) concludes that the role of emotional intelligence in developing one's writing fluency and relevancy is more than verbal intelligence. Pishghadam, Khodadady and Khoshsabk (2010) studied the impact of visual and verbal intelligences-based teaching on the vocabulary retention and written production of Iranian intermediate EFL learners. They found a significant difference in the visual experimental group but not in the verbal group. The relative consistency of the findings in these studies shows that the role of intelligence factors in language learning, particularly writing cannot be neglected.

**Higher-order Processes in L2 writing**

The study of higher-order cognitive processes in writing research started in the 70s. According to Stallard (1974), successful writers focus on content, organization, and audience and do not get lost in the midst of grammatical and spelling issues. In other words, advanced writers prioritize cognitive general factors or intelligence factors over linguistic competence factors particularly grammar knowledge (Bitchener & Knoch, 2010; Murphy & Roca de Larios, 2010). This is also supported by empirical research findings; Hall (1990) found that writers employ the same strategies and cognitive behaviors in L1 and L2. One of the most straightforward articulations of the significance of non-linguistic general cognitive factors in developing writing ability is found in Kobayashi and Rinnert (2008); the findings of this study provided evidence for the transferability of writing competence across languages.

Higher-order processes in the brain are not bound to the first or second language. The learners' organizational skills are transferable between L1 and L2. A great deal of knowledge now available on the dynamics of higher-order process in L2 writing comes from a range of transfer studies. Earlier the main focus of such studies was finding the manifestation of L1 elements in L2 products (Chen & Baker, 2010; Flowedew, 2010; Kenkel & Yates, 2009). However, for some scholars the concept of L1 use in L2 has a deeper dimension i.e. the reuse of L1 processes in the target language (De Larios, Marín, & Murphy, 2001; Sun-Alperin & Wang, 2011; Uzawa, 1996; Yamashita & Jiang, 2010). The process-oriented paradigm in writing research is concerned with cognitive behaviors that characterize the writing process (Pennington & So, 1993). Here "process" is synonymous to cognitive factor.

The higher-order processes governing both L1 and L2 proficiency are also addressed by Sparks and Gonschow (2001). According to their Linguistic Coding Differences Hypothesis (LCDH), linguistic coding works as a central cognitive factor which refers to L1 literacy skills including orthographic processing needed for writing tasks. According to Sparks and Gonschow (2001), such skills can predict L2 acquisition rate and proficiency to a considerable extent. Their findings show that successful L2 learners have stronger L1 literacy and syntactic skills. This is in accordance with the result of a longitudinal study conducted by Dufva and Voeten (1999) who examined L1 literacy acquisition and its impact of learning a foreign language. They concluded that the basis of L2 learning is partially formed by native language word recognition. The statistical associations found between L1 literacy and L2 writing might be caused by deeper cognitive factors which play a role in both languages. The investigation of the hypothetical role of cognitive factors common to

L1 and L2 literacy has formed a line of research in writing studies which will be discussed in the next section.

**Narrative Intelligence**

Narrative intelligence is defined by Randall (1999) as the ability to perceive and produce narrative structures. Compared to emotional and verbal intelligences whose roles in language learning have been studied during the past decade, narrative intelligence has received much less attention. In Randall's opinion, Gardner's (1983) theory of multiple intelligences has opened the door to other types of intelligence (other than Gardner's seven). He also proposes that narrative intelligence develops along with inter-personal, intra-personal, and verbal intelligences. Based on the theory of narrative intelligence, "We are all narratively intelligent to at least a minimal degree" (p. 15). The five dimensions of narrative intelligence include emplotment (creating the main structure and managing the general path of the events), characterization (producing a sufficiently elaborated account of the parties involved), narration (putting the events and characters in the right order from the beginning to the end), genre-ation (regarding the generic standards and reflecting them in one's narrative moves and general attitude), and thematization (reinforcing the message sent to the audience via using a system of signs enriched by the culture and knowledge shared by the writer and her audience). Each of these dimensions is then elaborated discussing their internal dynamics (see Randall, 1999).

Narrative intelligence of foreign language learners functions in both L1 and L2. However, measuring one's narrative ability must be done in L1 (not L2) to reduce possible error caused by the learners' insufficient L2 knowledge. That is to say, although narrative intelligence affects one's L2 performance, it should not be examined in L2. That is why the only validated scale of narrative intelligence (Pishghadam et al., 2011) which is based on Randall's (1999) theory was administered in the first language. Following the same logic, in L2 learning research higher-order processes (e.g. intelligence) are usually examined via learners' mother tongue (see Fahim & Pishghadam, 2007). Therefore, given the core idea of the present research i.e. including intelligence factors for explaining L2 writing ability, the scale devised by Pishghadam et al. (2011) seems to be the best option for measuring L2 writers' narrative intelligence.

## Method

**Participants**

Participants of the present study comprised 347 Iranian learners of English as a foreign language from four cities of Iran: Mashhad, Kashan, Lahijan and Tehran. The age of the participants ranged from 17 to 33. The sample included 268 university students majoring in English Language and Literature, Engineering, and Basic sciences, and the rest were high school students out of which 201 participants were females and 146 were males. All the participants were learners of English attending private English institutes (224 participants) or passing university ESP courses (123 participants). Each participant attended 6 test sessions. All the participants were informed about the general objectives of the project, gave their consent to participate

in the study and were assured of the confidentiality of any personal information they revealed during the study.

**Instrumentation**

The instruments used in this study include scales for measuring narrative intelligence, verbal intelligence, knowledge of grammar, depth and breadth of knowledge of vocabulary, and writing skill.

Pishghadam, Baghaei, Shams, and Shamsaee (2011) developed and validated an objective overall measure of narrative intelligence. They used Rasch analysis to substantiated the construct validity of the scale. This scale which includes 23 items assessing participants' performance on several dynamics of narrative intelligence (Randall, 1999) was used to measure participants' narrative intelligence in the present study. The scale includes 5 subsections that corresponds to five sub-abilities of narrative intelligence namely emplotment, characterization, narration, genre-ation, and thematization. The participants' ability for realizing each of the dynamics of narrative intelligence was rated separately and the total score indicated their narrative intelligence. The reliability (internal consistency) of this measure is 0.72 (Pishghadam et al., 2011). The inter-rater reliability of the scale was 0.83. The Alpha Cronbach for this instrument in the present study was 0.85.

To measure verbal intelligence of the subjects, the verbal scale of Wechsler's Adult Intelligence Scale (III) (1981) was used. The Farsi version of the WAIS Vocabulary subsection used in the present study consists of 40 words. This translated version was developed by Azmoon Padid institute (1993) in Tehran, Iran. The Alpha Cronbach for the vocabulary subsection in the present study was 0.68. The reliability coefficient (internal consistency) for the Verbal IQ is .97. The vocabulary subtest correlates highly (.91-.95) with the Verbal scale of the WAIS-III. The concurrent validity of WAIS-III is established based on high correlation with other valid intelligence scales. For example, "correlations between WAIS-III scores and Standford-Binet Intelligence Scale Fourth Edition (SB-IV) composite scores were high, ranging from 78 to 89" (Silva, 2008).

The structure module of TOEFL PBT published by ETS (2005) was used to measure participants` knowledge of English grammar. Since the validity of this scale had already been tested in the actual exam, the researchers found the scale appropriate to be used in the present study. This module contains 40 items. Fifteen items present a sentence with one part replaced by a blank. In the next 25 items, each sentence has four underlined words or phrases. It was required that the participants identify the wrong parts and mark them on the answer sheets. The Alpha Cronbach for this instrument in the present study was 0.80.

To measure the depth of participants' vocabulary knowledge, the Depth of Vocabulary (DVK) scale was used. The test contains 40 items. Each item consists of a stimulus word (adjectives) and eight choices. In each item, the first four choices (A-D) are in one box and the second four choices (E-H) are in another box. Among the choices of the left box, one to three choices could be synonymous to the stimulus, whereas among the four choices in the right box, one to three co-occurring words could be matched with the stimulus (collocations). The reliability of this test is

reported to be .91 (Qian, 1999). The Alpha Cronbach for this instrument in the present study was 0.76.

The second version of Vocabulary Levels Test (VLT) was used to measure the breadth of participants' vocabulary knowledge. The validity of the five sections of this test reported as Rasch ability estimates is as follows: 42.5 (2000), 45.9 (3000), 51.0 (5000), 55.2 (Academic), and 61.7 (10000). It measures the meaning of the content words via matching the definitions with the choices. For each three definitions, six choices are available, but each definition should be associated with only one choice. The measure is composed of five frequency levels (2000, 3000, 5000, academic, 10000) and thus is called the levels test. The first two levels (2000 and 3000) are composed of high frequency words. The 5000 level is considered a boundary level and the next two levels consist of words that generally appear in university texts (academic) and low frequency words (10000). The reliability of the different levels of this test was reported as follows; 2000 (.92); 3000 (.92); 5000 (.92); academic (.92); and 10000 (.96) (Schmitt et. al, 2001). The Alpha Cronbach for this instrument in the present study was 0.81. Schmitt et al. (2001) estimated the validity of the Levels Test by "establishing whether learners do better on the higher frequency sections than on the lower frequency ones" (p. 67). They found that out of 30 as the maximum, the mean for the frequency levels were as follows: 25.29 (sd 5.80) for the 2000 level, 21.39 (7.17) for the 3000 level, 18.66 (7.79) for the 5000 level and 9.34 (7.01) for the 10 000 level. According to them, analysis of variance plus Scheffe ´ tests showed that the differences were all statistically significant ($p < .001$). The validity of the Academic level section needs more explanation. The mean score of this section in the profile research done by Schmitt et al. (2001) was found to be 22.65 which apparently places it somewhere between the 2000 level and 3000 level. However, they argue that the words in this section are different from the other levels, and therefore should not be included in the profile comparison. The validity of this section is then justified by analyzing the facility values of individual items and Rasch item difficulty measures. According to Schmitt et al. (2001), "the figures suggest that the words in the academic level fit in a broad range between the 2000 level and the 10 000 level" (p. 68).

To measure the participants writing ability, the researchers used an original specimen of the writing module of the IELTS exam published by ETS (2005) whose validity had been already substantiated by ETS. Half-band scores were included. Task 2 of the General Training Writing Module was assessed based on 1) coherence and cohesion; 2) lexical resource; and 3) grammatical range and accuracy. The task requires the candidates to formulate and develop a position in relation to a given prompt in the form of a question or statement. The inter-rater reliability of the scale was 0.87.

**Procedure**

The samples were gathered across the five cities used as the sampling pool. Other than the narrative intelligence test which was administered via a movie session and recording participants' voice, the other five tests were given to them in traditional setting of paper and pencil exams. At the first phase of the study, the participants took the writing test and their performance was rated based on IELTS scoring

criteria. This produced a set of writing scores on a scale of 1 to 9 with half-band scores. Then, the test of grammar was taken by participants and each person received a score out of 40. In the next step, the depth of vocabulary test was administered and the participants were asked to mark four choices altogether for each item. This test produced a set of scores ranging from 0 to 100. Then the depth of vocabulary test was given to the participants. The participants' scores on this test were given on a scale of 0 to 160. After that the Verbal Intelligence Test was administered during which each participant was presented with 1 word at a time and asked to explain each word's meaning verbally. The examiner rates the responses with a 0, 1, or 2 depending on how well the participant defines the word. Therefore, the scores can range from 0 to 80 (Wechsler, 1997). The last phase was the administration of the narrative intelligence test. The participants watched the first 10 minutes of a movie (*Defiance*) and then, were asked to recount the story. They were also asked to tell their story of the first day of the elementary school. The two narratives produced by each participant were then rated by two raters using the NIS (Narrative Intelligence Scale). The average score for the five sub-abilities of narrative intelligence in the above narrative tasks were taken as the participants' narrative intelligence score.

First of all, the internal reliability of the tests used in the study was calculated using the Alpha Cronbach Method. After ensuring the reliability of the scores, all the data were imported into SPSS 18.0 and linked to AMOS 16.0 to be analyzed through structural equation modeling (SEM). The observed variables in the models represent the collected data and the latent variables represent the hypothetical constructs which are assumed to play a role in developing learners' writing ability. Two models, one including only linguistic factors and the other one including intelligence factors as well, were linked to the data and their fit indices and parameter estimates were calculated by AMOS.

The use of structural equation modeling in the present study can be justified from two perspectives. First, the analytic solidarity found in SEM which is originated in its ability to process simultaneous equations including a range of dynamic variables (variables which play the role dependent and independent factor intermittently) exceeds that of others including regression analysis, path analysis and factor analysis. Actually, the fact that SEM is much less frequently used in applied linguistics studies compared to the mentioned types of analysis does not mean that those analyses are better than SEM; it is the complexity of data analysis in SEM from which researchers usually evade.

The second reason for using structural equation modeling in this study is the inclusion of latent variables in SEM models which can provide the researchers with the opportunity to test their hypotheses about the assumed constructs which cannot be directly measured. Adding latent variables, in fact, is an attempt to make the prediction models in social sciences more realistic since researchers know that they cannot measure the constructs straightly and have to resort to measuring participants' performance which is affected by various factors including error factors. Therefore, two types of latent variables are included in SEM models: error variables and latent constructs affecting the scores obtained by the participants. The rest of the variables are all observed variables. SEM models include two sections: the

measurement model and the structure models. The measurement model relates observed variables to latent variables and the structure model relates latent variables to each other. The combination of these two models creates a range of simultaneous equations which are saturated using the data presented to the SEM model.

The two SEM models used in this study present two different combinations of observed and latent variables. The next section introduces the models and their justifications based on the literature of writing research and cognitive sciences.

## Results and Discussion

In the present study, six sets of data were collected through the administration of several tests. The descriptive statistics of the scores obtained by all 347 participants on these tests is presented in Table 1.

Table 1
*The Descriptive Stat. of the Six Tests Administered in the Study*

| | Mean | Std. Deviation | Std. Error of Measurement | Min. | Max. |
|---|---|---|---|---|---|
| ***Grammar*** | 57.51 | 16.17 | 0.89 | 23 | 98 |
| ***Depth of Vocabulary*** | 41.04 | 14.19 | 0.76 | 7 | 88 |
| ***Breadth of Vocabulary*** | 44.54 | 18.91 | 1.01 | 12 | 100 |
| ***Verbal Intelligence*** | 73.20 | 6.91 | 0.37 | 54 | 93 |
| ***Narrative Intelligence*** | 56.07 | 10.09 | 0.54 | 36 | 90 |
| ***Writing*** | 43.56 | 13.17 | 0.70 | 17 | 89 |

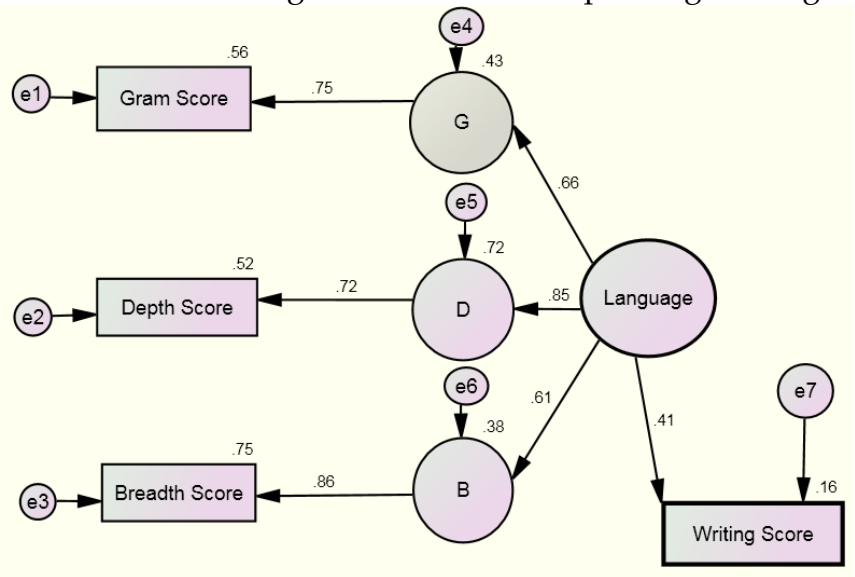
As Table 1 shows, verbal intelligence has the highest mean among the other constructs while depth of vocabulary has the minimum mean value. It should be mentioned that these tests were administered with different rating scales; here for the sake of homogeneity all of the scales are converted to a 0 to 100 scale so that comparisons can be made more easily. The Std. Deviations of the scores show that participants' verbal intelligence is the most homogeneous construct while the most heterogeneity is observed in breadth of vocabulary with a Std. deviation of more than 18. The widest range of scores belongs to breadth of vocabulary and the narrowest one belongs to verbal intelligence.

### SEM Parameter Estimates

Model 1 represents the view based on which only linguistic factors determine one's writing ability in a foreign language. Such a view has been supported by Jeyaraj (2010) and Coxhead and Byrd (2007). In the literature these factors are labeled in various ways. Based on the literature *grammar knowledge* (Andrews et al., 2006; Mair, 2007), *depth of vocabulary knowledge* (Chang, Chang, Chen, & Liou, 2008; Laufer & Waldman, 2011) and *breadth of vocabulary knowledge* (Lee, 2003; Stæhr, 2008; Webb, 2009) are the three main linguistic factors which can determine one's writing ability in L2. Therefore, Model 1 which represents this view only includes these factors as the predictors of L2 writing. The existence and direction of the arrows in the model reflects the assumptions held by the above-mentioned scholars according to which

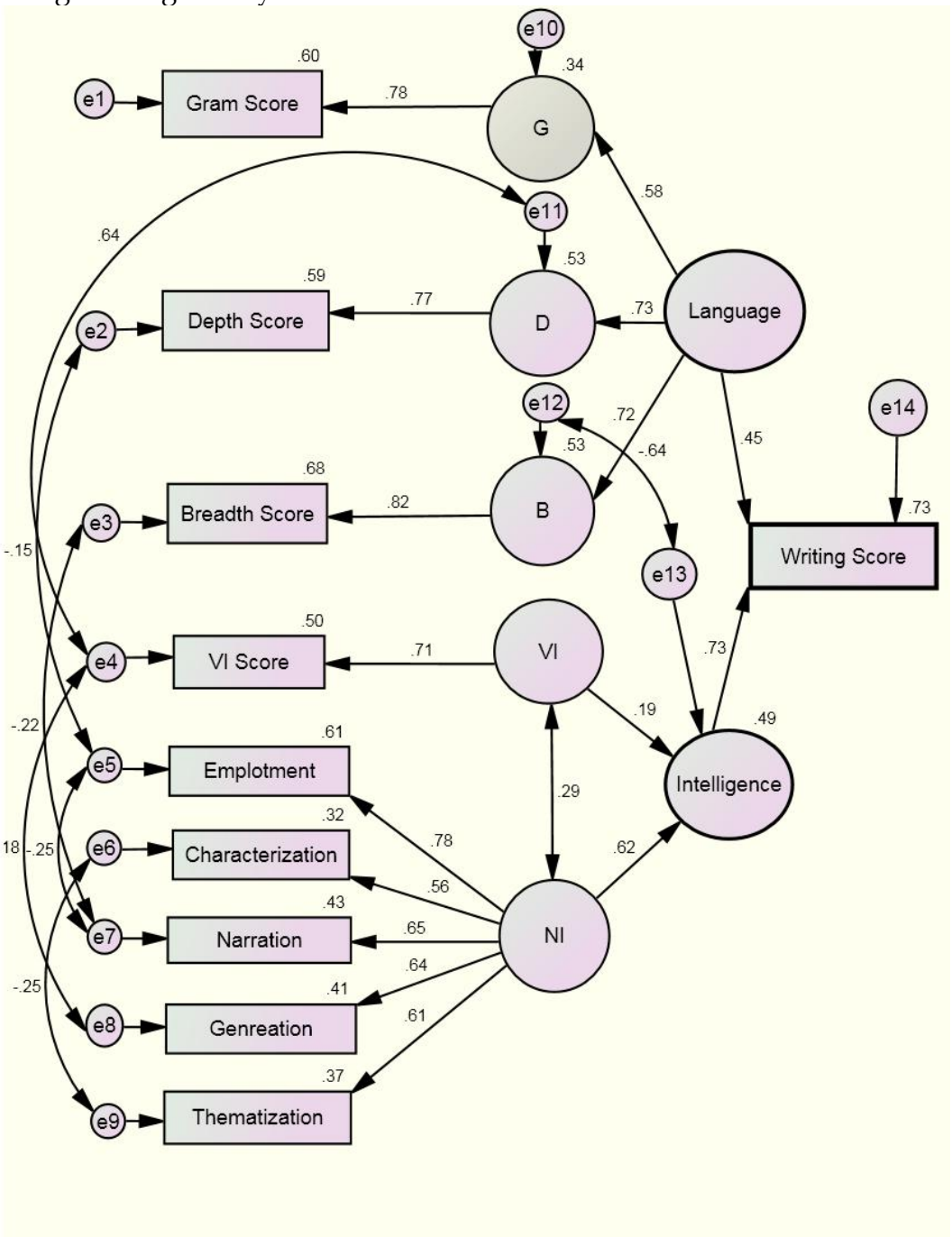
linguistic competencies comprising the language faculty in human mind only include Grammar (G), Depth of vocabulary (D) which is usually referred to as learners' knowledge of target language collocation, and breadth of vocabulary (B). Each of these sub-constructs produces an observed score which is also affected by unknown factors altogether labeled as e1 in the model.

*Figure 1.* Model 1 with Three Linguistic Factors for Explaining Writing Ability




The path numbers show standardized estimates of the model parameters (correlation and regression coefficients). The numbers shown above rectangles (observed variables) and circles (latent variables) show the variation explained by the paths leading to it. According to this model and the gathered data, among the three sub-factors of linguistic competence as predictors of writing ability, the scores obtained by the learners for breadth of vocabulary knowledge are associated with less measuring error; three fourth (%75) of the variation in the breadth scores can be accounted for by linguistic competence. The explained variance of depth (%52) and grammar (%56) scores are close. Altogether, all the linguistic factors can explain only %16 of the variance observed in the writing scores. According to Model 1, the other %84 of the variance cannot be accounted. The researchers' hypothesis is that a considerable part of the unexplained variance of the writing ability in Model 1 can be accounted for by verbal and narrative intelligence factors. In accordance with this hypothesis, the intelligence factors are incorporated into Model 2 which is shown below in Figure 2.

*Figure 2.* Model 2 with Two Intelligence Factors and Three Linguistic Factors for Explaining Writing Ability




SEM Model 2, which is presented here for the first time, indicates the researchers' view on the cognitive factors that play a meaningful role in developing language learners` writing ability. In model 2, linguistic competence is demonstrated through grammar knowledge, depth of vocabulary knowledge, and breadth of vocabulary knowledge while intelligence develops out of verbal intelligence and narrative intelligence. The necessity of including cognitive factors in a model of writing has been implicitly and explicitly supported by Bourke and Adams (2010), Cavanagh and Langevin (2010), Gustilo (2010), and Lee and Tan (2010) but so far no attempt has been made to put their claims into test. Moreover, the inclusion of intelligence factors is occasionally suggested in the literature (e.g. see Abiodun &

Folaranmi, 2007; Dobson, 2005; Hussein, 2008; Pishghadam, 2009) but has rarely been statistically studied to date.

Verbal and narrative intelligences are two sub-factors which are added to Model 1 to improve the fitness. In Model 1, only %16 of the variance of writing scores can be explained by the independent variables of the study. Adding intelligence factors improved this parameter by %47 which is quite significant. The latent variable "intelligence" in Model 2 develops out of two other latent variables "verbal intelligence" and "narrative intelligence" which altogether can explain %49 of the variation observed in the participants' "intelligence"; Of course "narrative intelligence" is much more explanatory than "verbal intelligence" (0.62>0.19). Among the observed scores for the five sub-abilities of narrative intelligence, "emplotment" scores show the highest variance explained by the latent variable "narrative intelligence" in Model 2; this variable can also explain "narration" (%43), "genre-ation" (%41), "thematization" (%37), and "characterizarion" (%32) with respective degrees of explanatory power.

**SEM Fit Indices**

If the fitting indices of Model 2 (proposed by the researchers) are better than Model 1 (based on the current beliefs about writing ability) then the hypothesis is corroborated. In other words, if Model 2 (including intelligence factors) fits the collected data better than Model 1(lacking intelligence factors), one can argue that a theory of foreign language learners' writing ability which considers the role of intelligences, specially narrative intelligence, can explain the relationship between the variables involved in writing better than a theory than excludes intelligence factors. Each of the fit indices in structural equation modeling has an acceptable range. For doing the comparison between competing SEM models, the values which are within the acceptable range of fit can be used to compare several models. The fitting cut-off values in the present study are adopted from the recent SEM references (e.g. see Kaplan, 2009). Given a number of features such as sample size, normality of gathered data, and the nature of variables involved in each research project a certain set of absolute and relative fit indices are usually selected and reported. In the present study the following fit indices are used:

1. $\chi^2/df$ **:** it is the ratio of chi-square value to the model's degree of freedom. The chi-square tests the hypothesis that the model perfectly fits the data.
2. AGFI (Adjusted Goodness of Fit Index): it takes into account the model's degree of freedom. (Arbuckle, 2007)
3. IFI (Incremental Fit Index): it compares model's degree of freedom and discrepancy to those of the baseline model. (Arbuckle, 2007)
4. TLI (Tucker-Lewis Index): it depends on the correlation among the variables in the model; it is used to compare competing models. (Fornell & Larcker, 1981)
5. CFI (Comparative Fit Index): it is similar to TLI. In addition, it considers the increment in non-centrality. (Schmacker & Lomax, 2004)

6. RMSEA (Root Mean Square of Approximation): it shows the badness of fit. The lower it is, the more evidence exist that the models fit the data. It is usually used for comparing two competing models. (Schmacker & Lomax, 2004)

The fitting indices for Model 1 (without intelligence factors) and Model 2 (with intelligence factors) are shown in Table 2.

Table 2
*Fitting Indices for Model 1 (Excluding Intelligence) and Model 2 (Including Intelligence)*

| Fit Index | $\chi^2/df$ | *AGFI* | *IFI* | *TLI* | *CFI* | *RMSEA* |
|---|---|---|---|---|---|---|
| Acceptable Range | < 3 | > 90 | | | | < 0.08 |
| Model 1 | **8.05** | **0.89** | **0.84** | **0.67** | **0.83** | **0.14** |
| Model 2 | **1.98** | **0.94** | **0.96** | **0.93** | **0.96** | **0.05** |

As it can be seen, Model 1 does not have a good fit while Model 2 does. The Chi-square of Model 2 (1.98) is within the fitting range while Model 1's (8.05) is not. The main index for fitting the data is AGFI; according to this index Model 1 (0.89) is slightly below the acceptable range (> 90) while Model 2 (0.94) is in the safe area. IFI index shows the same pattern with a bigger difference between the models (Model 1: 0.84; Model 2: 0.96). TLI which is specifically designed for comparing competing models shows a considerable distance between the models; Model 2 (0.93) is superior to Model 1 (0.67). CFI shows the same pattern with a smaller distance though. Last but not least is the RMSEA index which shows the badness of fit and is a reliable index for comparing the competing models. The RMSEA of Model 1 is well beyond the fitting range while Model 2's is small enough to be acceptable. All in all, this means that a model including verbal intelligence and narrative intelligence as cognitive predictors of writing ability can explain the data better than a model that excludes those factors and only relies on linguistic factors as predictors of writing. In addition, Model 2 is superior to Model 1 in that the parameter estimates of the two main latent variables predict the variance observed in the writing scores up to %73 which is way more than %16 which the amount of variance explained in Model 1. The inclusion of verbal and narrative intelligences has clearly increased the explanatory power of the latent predictors. In Model 2 only %27 of the variance observed in the writing scores is not accounted for.

## Conclusion

The main goal of this study was to examine to what extent linguistic and non-linguistic factors can account for the writing ability of the foreign language learners. To this end, two models were proposed by the researchers. In Model 1, only linguistic factors (grammar, depth and breadth of vocabulary) were taken into consideration while in Model 2 linguistic and intelligence factors were proposed as predictors of L2 writing ability. As it was found in the present study, L2 writers' knowledge of vocabulary and grammar can only account for 16 percent of variance observed in their writing ability. The findings also exhibited that including verbal

and narrative intelligences as general cognitive abilities can increase the explained variance from 16 to 73 percent. The results also demonstrated that narrative intelligence more than verbal intelligence accounts for variance in L2 writing ability. It implies that narrative intelligence can fill the wide gap in L2 writing research and partially but sufficiently addresses this problem: "why do learners with the same knowledge of grammar and vocabulary have variant writing abilities?" The answer provided by Model 2 is this: "because they have different narrative intelligence levels." Therefore, it is the presence of narrative intelligence along with verbal intelligence in Model 2 that boosts the amount of accounted variance observed in the writing scores. This shows that macro non-linguistic factors do play an important and undeniable role in L2 writing.

Logically, foreign language learners' L2 performance must be affected by their verbal intelligence. Tests of verbal intelligence on the surface look like vocabulary scales; in such tests, participants' linguistic perception and production is examined via rating their choice of word, the brevity and sufficiency of the provided definitions, and their ability to express their ideas (see Wechsler, 1997). Tests of verbal intelligence are oral. In the writing mode, the mentioned productive skills will be reflected in the participants' lexical resources, the observed grammatical range and accuracy, and their ability to express their ideas via written discourse. The rating criteria for L2 writing exams particularly the criteria for rating candidates' lexical resources in IELTS (see ETS, 2005) imply that candidates with a higher verbal intelligence can use their lexical resources better. Therefore verbal intelligence could be considered as a viable option as one of the factors in an explanatory model of L2 writing. The findings of this study show that verbal intelligence explains only 19 percent of the variance observed in intelligence factors. Although this is a smaller percentage compared to that of narrative intelligence, it still shows the role of verbal intelligence in L2 writing.

Adding intelligence factors to a model of writing factors gives one a more fitting grasp of the real second language writing experience. Logically, the relatively large variance observed in writing scores observed in Model 1 must have been created either by error or by other latent variables which are independent of linguistic factors which can affect the writing ability to a considerable extent. It seems that writing needs more than the knowledge of vocabulary and grammar. The researchers believe that identifying the role of intelligence factors in writing is only the first step. To translate this knowledge into useful pedagogical methods, one has to move up to an explanatory phase of analysis: how do verbal and narrative intelligences contribute to L2 writing ability? To answer this question, the possible links between the dynamics of these cognitive factors and their counterparts in the literature of writing research have to be discussed.

Narrative intelligence, as a higher-order process, can contribute to the organizational skills of foreign language writers. Dynamics of narrative intelligence as defined by Randall (1999) and operationalized by Pishghadam et al. (2011) are comparable to a number of the higher-order processes discussed by writing researchers particularly organizational skills. Randall's theory of narrative intelligence tries to explain the cognitive processes involved in the coherent expression of ideas through language. The main objective in a writing exam is to

examine candidates' ability in putting their ideas into written discourse as coherently as possible.

Coherence of writing can be successfully maintained if the writer is narratively intelligent. A coherent piece of writing must have a good organization. In other words, they are good planners and maintain the fluency and coherence of the written discourse (maintain the central line of argument) by the appropriate use of connectives and logical links. These organizing skills discussed in Randall's (1999) work are considered as the dynamics of emplotment which can be defined as the ability to explain events in terms of origins, outcomes, influences and results. In the context of a writing exam, this means that the candidates should be able to introduce and discuss their ideas while maintaining the logical links between the sentences and paragraphs. Writers must create, select, and assign roles to the characters (not necessarily persons) in their writing; in other words, an important part of the content of writing is formed via characterization. With good narration the writers can "arrange recounted events with the right rhythm and ethos" (Randall, 1999, p.23); in genre-ation, the writer attends to the writing moves and general mood of the writing the learners are supposed to maintain. When the general mood is not steady, writer's attitude cannot be inferred from his writing and it cannot communicate the intended massage. Good writers always keep track of the main theme and do not digress. Thematization helps the learners link the paragraphs to each other and also bind the sentences within the paragraphs to maintain the integrity of their writing. Successful writers are good at planning (emplotment), presenting the concepts (characterization and narration), and maintaining the logical flow of ideas (genre-ation and thematization) in their writing. The results of the SEM modeling in this study show that the above-mentioned links between good writing and high narrative intelligence are statistically significant.

Teachers can improve L2 learners' writing ability through narrative intervention programs. This has been partially recognized by few scholars; however, a firm theoretical ground is needed to provide the teachers with enough momentum to develop their narrative intervention programs in L2 writing classrooms. Narrative literacy plays an important role in L2 writing programs. The results of this study can contribute to the pedagogy and assessment of second language writing. If including intelligence factors in a model of L2 writing projects a more realistic image of the reality of language learning, excluding them from writing classroom and writing assessment frameworks would not be a viable option. Narrative intervention programs set to surge L2 writers' narrative intelligence can help them with improving organization, content, and fluency of their writings. It can be suggested that focus on organization, content, and audience along with the dynamics of narrative intelligence be prioritized over syntactic and lexical concerns in writing. In addition, writing programs might benefit from intervention agendas which aim to promote learners' narrative competence. Being aware of the role of non-linguistic factors in developing candidates' writing ability, the designers of high-stake tests of English such as IELTS may need to redefine the assessment criteria of the writing module.

The findings of the present study generated new questions for the study of L2 writing and intelligence factors which need to be addressed in future research. The

inclusion of cognitive factors in a model of writing implies the necessity of interdisciplinary study of this complex skill. Further research can pursue investigating the interrelationship of the dynamics of narrative intelligence and discourse features of learner`s written corpora, exploring the cognitive and metacognitive processes in L2 which correlate with high verbal and narrative intelligence, studying the interaction of linguistic and cognitive factors for predicting the writing ability, using experimental designs to test the practical value of intelligence-informed teaching agendas, and designing research projects for testing the neuropsychological validity of the role of intelligence factor in L2 writing.

**References**

Abiodun, A. J., & Folaranmi, A. O. (2007). Effects of verbal ability on second language writers' achievement in essay writing in English language. *International Journal of African & African American Studies, 6*, 61-67.

Andrews, R., Torgerson, C., Beverton, S., Freeman, A., Locke, T., Low, G., Robinson, A., & Zhu, D. (2006). The effect of grammar teaching on writing development. *British Educational Research Journal, 32*, 39-55.

Arbuckle, J. A. (2007). *Amos 16.0 user's guide.* Chicago: AMOS Development Corporation.

Azmoon Padid Institute. (1993). *The standardization of Wechsler's adult intelligence scale III*. Tehran, Iran.

Barnes, M., Dennis, M., & Hetherington, R. (2004). Reading and writing skills in young adults with Spina Bifida and Hydrocephalus. *Journal of International Neuropsychological Society, 10*, 655-663.

Berman, R., & Cheng, L. (2001). English academic language skills: Perceived difficulties by undergraduate and graduate students, and their academic achievement. *Canadian Journal of Applied Linguistics, 4*, 25–40.

Bialystok, E. (2002). Acquisition of literacy in bilingual children: A framework for research. *Language learning, 52*, 159-199.

Bitchener, J., & Knoch, U. (2010). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing, 19*(4), 207-217.

Bloome, D., Katz, L., & Champion, T. (2003). Young children's narratives and ideologies of language in classrooms. *Reading & Writing Quarterly, 19*, 205-223.

Bourke, L., & Adams, A. M. (2010). Cognitive constraints and the early learning goals in writing. *Journal of Research in Reading, 33*, 94-110.

British Council. (2005). *IELTS specimen materials.* Cambridge: UCLES.

Cavanagh, M. O., & Langevin, R. (2010). The quality of French minority students' fictional texts: A study of the influence of a preferential cognitive style and writing strategy scaffolding. *Language, Culture and Curriculum, 23*, 71-87.

Chang, Y. C., Chang, J. S., Chen, H. J., & Liou, H. C. (2008). An automatic collocation writing assistant for Taiwanese EFL learners: A case of corpus-based NLP technology. *Computer Assisted Language Learning, 21*, 283-299.

Chen, Y. H., & Baker, P. (2010). Lexical bundles in l1 and l2 academic writing. Language *Learning and Technology, 14*, 30-49.

Chien, S. C. (2011). Discourse organization in high school students' writing and their teachers' writing instruction: The case of Taiwan. *Foreign Language Annals, 44*, 417-435.

Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing, 16*(3), 129-147.

Crook, C., & Bennett, L. (2007). Does using a computer disturb the organization of children's writing? *British Journal of Developmental Psychology, 25*, 313-321.

De Larios, J. R., Marín, J., & Murphy, L. (2001). A temporal analysis of formulation processes in L1 and L2 writing. *Language Learning, 51*, 497-538.

Devine, J., Railey, K., & Boshoff, P. (1993). The implications of cognitive models for in L1 and L2 writing. *Journal of Second Language Writing, 2*, 223-235.

Dobson, S. (2005). Narrative competence and the enhancement of literacy: Some theoretical reflections. *International Journal of Media, Technology, and Lifelong Learning, 1*(2), 1-14.

Dufva, M., & Voeten, M. (1999). Native language literacy and phonological memory as prerequisites for learning English as a foreign language. *Applied Psycholinguistics, 20*, 329-348.

Duong, M. T., Cuc, N. T. K., & Griffin, P. (2011). Developing a framework to measure process-oriented writing competence: A case of Vietnamese EFL students' formal portfolio assessment. *RELC Journal 42*(2), 167-185.

Elbow, P. (2006). The music of form: Rethinking organization in writing. *College Composition and Communication, 57*, 620-666.

Eng, L. L., & Mustapha, G. (2010). Enhancing writing ability through multiple-intelligence strategies. *Journal of Social Science and Humanities, 18*, 53-63

Erkan, D. Y., & Saban, A. I. (2011). Writing performance relative to writing apprehension, self-efficacy in writing, and attitudes towards writing: A correlational study in Turkish tertiary-level EFL. *Asian EFL Journal, 12*(1), 164-192.

Escribano, P. D. (1999). Teaching writing through reading: A text-centered approach. *IBÉRICA, 1*, 55-62.

ETS. (2005). *IELTS specimen material.* Cambridge: Cambridge University Press.

ETS. (2005). *TOEFL actual tests: 7 full-length paper-based exams administered in the past by ETS.* Cambridge: Cambridge University Press.

Fahim, M., & Pishghadam, R. (2007). On the role of emotional, psychometric, and verbal intelligences in the academic achievement of university students majoring in English language. *Asian EFL Journal, 9*, 240-253.

Ferris, D. R. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing, 8*, 1–10.

Ferris, D. R. (2004). The ''grammar correction'' debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime . . .?). *The Journal of Second Language Writing, 13*, 49-62.

Ferris, D. R. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. *Studies in Second Language Acquisition, 32*(2), 181-201.

Flowerdew, J. (2010). Use of signaling nouns across L1 and L2 writer corpora. *International Journal of Corpus Linguistics, 15*, 36-55.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 48*, 39–50.

Gardner, H. (1983). *Frames of mind.* New York: Basic books.

Goldberg, T., Schwarz, B. B., & Porat, D. (2011). Could they do it differently? Narrative and argumentative changes in students' writing following discussion of hot historical issues. *Cognition and Instruction, 29*, 185-217.

Gupta, D., & Woldemariam, G. S. (2011). The influence of motivation and attitude on writing strategy use of undergraduate EFL students: Quantitative and qualitative perspectives. *Asian EFL Journal, 13*(2), 34-89.

Gustilo, L. E. (2010). What you think, feel, and experience shape your writing: Cognitive, affective, and contextual processes in ESL writing. *Asia-Pacific Education Researcher, 19*, 271-285.

Hamp-Lyons, L., & Mathias, S. P. (1994). Examining expert judgments of task difficulty on essay tests. *Journal of Second Language Writing, 3*, 49–68.

Hirose, K. (2006). Pursuing the complexity of the relationship between L1 and L2 writing. *Journal of Second Language Writing, 15*, 142-146.

Hussein, J. W. (2008). An existential approach to engaging adult learners in the process of legitimizing and constructing meanings from their narrative knowledge. *Action Research, 6*, 391-420.

Hyland, K. (2008). 'Small bits of textual material': A discourse analysis of Swales' writing. *English for Specific Purposes, 27*, 143-160.

Jarvis, S., Grant, L., Bikowski, D., & Ferris, D. (2003). Exploring multiple profiles if highly rated learner compositions. *Journal of Second Language Writing, 12*, 377-403.

Jeyaraj, J. (2010). British Indian grammar, writing pedagogies, and writing for the professions: Classical pedagogy in British India. *Technical Communication Quarterly, 19*, 379-402.

Kaplan, D. (2009). *Structural equation modeling.* London: Sage Publication.

Kenkel, J., & Yates, R. (2009). The Interlanguage grammar of information management in L1 and L2 developing writing. *Written Communication, 26*, 392-416.

Kobayashi, H., & Rinnert, C. (2008). Task response and text construction across L1 and L2 writing. *Journal of Second Language Writing, 17*, 7-29.

Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing, 20*, 148-161.

Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing, 17*, 48-60.

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning, 61*, 647-672.

Lee, C. C., & Tan, S. C. (2010). Scaffolding writing using feedback in students' graphic organizers: Novice writers' relevance of ideas and cognitive loads. *Educational Media International, 47*(2), 135-152.

Lee, S. H. (2003). ESL learners' vocabulary use in writing and the effects of explicit vocabulary instruction. *System, 31*, 537-561.

Lei, X. (2008). Exploring a sociocultural approach to writing strategy research: Mediated actions in writing activities. *Journal of Second Language Writing, 17*(4), 217-236.

Leijten, M., van Waes, L., & Ransdell, S. (2010). Correcting text production errors: Isolating the effects of writing mode from error span, input mode, and lexicality. *Written Communication, 27*, 189-227.

Lowry, P. B., Curtis, A., & Lowry, M. R. (2004). Building a taxonomy and nomenclature of collaborative writing to improve interdisciplinary research and practice. *Journal of Business Communication, 41*, 66-99.

Magnifico, A. M. (2010). Writing for whom? Cognition, motivation, and a writer's audience. *Educational Psychologist, 45*(3), 167-184.

Mair, C. (2007). British English/American English grammar: Convergence in writing - Divergence in speech? *Anglia, 125*, 84-100.

Martinez, C. T., Kock, N., & Cass, J. (2011). Pain and pleasure in short essay writing: Factors predicting university students writing anxiety and writing self-efficacy. *Journal of Adolescent and Adult Literacy, 54*, 351-360.

Murphy, L., & Roca de Larios, J. (2010). Searching for words: One strategic use of the mother tongue by advanced Spanish EFL writers. *Journal of Second Language Writing, 19*(2), 61-81.

Myhill, D., Jones, S. (2007). More than just error correction: Students' perspectives on their revision processes during writing. *Written Communication, 24*, 323-343.

Nakamaru, S. (2010). Lexical issues in writing center tutorials with international and US-educated multilingual writers. *The Journal of Second Language Writing, 19*, 95-113.

Olinghouse, N. G., & Leaird, J. T. (2009). The relationship between measures of vocabulary and narrative writing quality in second- and fourth-grade students. *Reading and Writing, 22*, 545-565.

Ong, J., & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing, 19*, 218-233.

Pantaleo, S. (2010). Developing narrative competence through reading and writing metafictive texts. *Literacy Research and Instruction, 49*, 264-281.

Paradis, M., & Hildebrandt, N. (1985). *Neurolinguistic aspects of the Japanese writing system*. Florida: Academic Press.

Pennington, M. C., & So, S. (1993). Comparing writing process and product across two languages: A study of 6 Singaporean university student writers. *Journal of Second Language Writing, 2*, 41-63.

Pishghadam, R. (2009). Emotional and verbal intelligences in language learning. *Iranian Journal of Language Studies, 3*(1), 43-64.

Pishghadam, R., & Ghonsooly, B. (2008). On the role of emotional intelligence in second language learning success. *Pazhuheshe Zabanhaye Khareji, 43*, 47-59.

Pishghadam, R., & Ghonsooly, B. (2008). On the role of emotional intelligence in second language learning success. *Pazhuheshe Zabanhaye Khareji, 43*, 47-59.

Pishghadam, R., Baghaei, P., Shams, M., & Shamsaee, S. (2011). Construction and validation of a narrative intelligence scale with the Rasch rating scale model. *International Journal of Educational and Psychological Assessment, 8*(1), 75-90.

Pishghadam, R., Khodadady, E., & Khoshsabk, N. (2010). The impact of visual and verbal intelligences-based teaching on the vocabulary retention and written production of Iranian intermediate EFL learners. *MJAL, 2*(5), 379-395.

Rahimi, A., & Qannadzadeh, J. A. (2010). Quantitative usage of logical connectors in Iranians' EFL Essay writing and logical and linguistic intelligences. *Social and Behavioral Sciences, 5*, 2012-2019.

Qian, D.D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review, 56*, 282-308.

Randall, W. L. (1999). Narrative intelligence and the novelty of our lives. *Journal of Aging Studies, 13*(1), 11-28.

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

Sandberg, A. D., & Hjelmquist, E. (1996). A comparative, descriptive study of reading and writing skills among non-speaking children: A preliminary study. *International Journal of Language & Communication Disorders, 31*, 289–308.

Schmitt, N., Schmitt, D., & Calpham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test**.** *Language Testing, 18*, 55-88.

Schneider, C. (2011). Why field linguists should pay more attention to research in applied linguistics. *Australian Journal of Linguistics, 31*(2), pp. 187-209.

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling.* London: Lawrence Erlbaum Association.

Shanahan, T. (2006). Relations among oral language, reading, and writing development. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 171-186). New York, NY: Guilford Press.

Silva, M. A. (2008). Development of the WAIS-III: A brief overview, history, and Description. *Graduate Journal of Counseling Psychology, 1*, 117-135.

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183–205). Cambridge: Cambridge University Press.

Snider, R. C. (2002). *The effectiveness of oral expression through the use of continuous speech recognition technology in supporting the written composition of postsecondary students with learning disabilities.* Unpublished doctoral dissertation, Virginia Polytechnic Institute.

Sparks, R., & Gonschow, L. (2001). Aptitude for learning a foreign language. *Annual Review of Applied Linguistics, 21*, 90-111.

Stæhr, L.S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal, 36*, 139-152.

Sun-Alperin, M. K., & Wang, M. (2011). Cross-language transfer of phonological and orthographic processing skills from Spanish L1 to English L2. *Reading and Writing, 24*, 591-614.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46*, 327-369.

Truscott, J. (1999). The case for ''the case for grammar correction in L2 writing classes'': A response to Ferris. *Journal of Second Language Writing, 8*, 111-122.

Uttl, B., & Van Alstine, C. L. (2003). Rising verbal intelligence scores: Implications for research and clinical practice. *Psychology and Aging, 18*, 616-621.

Uzawa, K. (1996). Second language learners' processes of L1 writing, L2 writing, and translation from L1 into L2. *Journal of Second Language Writing, 5*, 271-294.

Vaid, J., & Padakannaya, P. (2004). Reading and writing: An interdisciplinary journal. *Reading and Writing, 17*, 1-6.

Webb, S. A. (2009). The effects of pre-learning vocabulary on reading comprehension and writing. *Canadian Modern Language Review, 65*, 441-470.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale Revised.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997). *WAIS–III administration and scoring manual*. San Antonio, TX: Psychological Corporation.

Weekes, B. S., Yin, W., Su, I. F., & Chen, M. J. (2006). The cognitive neuropsychology of reading and writing in Chinese. *Language and Linguistics, 7*, 595-617.

Widdowson, H. G. (2000). On the limitations of linguistics applied. *Applied Linguistics, 21*, 3-5.

Wolsey, T. D. (2010). Complexity in student writing: The relationship between the task and vocabulary uptake. *Literacy Research and Instruction, 49*(2), 194-208.

Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of l2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly, 44*, 647-668.

Zhou, A. A. (2009). What adult ESL learners say about improving grammar and vocabulary in their writing for academic purposes? *Language Awareness, 18*, 31-46.

# Students' Perceptions of the Impact of the College English Test

**HONGLI LI**
*Georgia State University, USA*

**QI ZHONG**
*Central University of Finance and Economics, China*

**HOI K. SUEN**
*The Pennsylvania State University, USA*

**Bio Data:**
Hongli Li is an assistant professor of Research, Measurement, and Statistics in the Department of Educational Policy Studies at Georgia State University, USA. Her research interests include the use of testing to inform teaching and learning, diagnostic assessment of reading comprehension, and language testing.

Qi Zhong is a lecturer with the School of Foreign Languages at the Central University of Finance and Economics, China. Her research interests include teaching English as a second language, comparative literature and culture, and business English speech.

Hoi K. Suen is a distinguished professor of Educational Psychology at The Pennsylvania State University, USA. His research interests include measurement theories, behavioral observation, validity issues, consequences of high-stakes testing, and lessons from the historical civil service exam system of China.

**Abstract**
Designed to assess college students' English ability, the College English Test (CET) is regarded as the most influential English test in China. This study investigates students' perceptions of the impact of the CET on their English-learning practices and their affective conditions. A survey was administered to 150 undergraduate students at a university in Beijing. It was found that students perceived the impact of the CET to be pervasive. In particular, the majority of the respondents indicated that the CET had a greater impact on what they studied than on how they studied. Most of the students surveyed felt the CET had motivated them to make a greater effort to learn English. Many students seemed to be willing to put more effort on the language skills most heavily weighted in the CET. About half of the students reported

a higher level of self-efficacy in regard to their overall English ability and some specific English skills as a result of taking or preparing for the CET. However, many students also reported experiencing increased pressure and anxiety in relation to learning English. This study provides important evidence about how the CET influences college students' English learning in China, and directions for further research are also suggested.

*Keywords:* Students, Perceptions, Impact, College English Test

## Introduction

Launched by the Ministry of Education, People's Republic of China, in 1987, the College English Test (CET) is a standardized test designed to measure the English proficiency of undergraduate students in China and to determine whether their English-language ability meets the requirements of the national college English curriculum. In accordance with the national curriculum, the CET consists of two tests: Band 4 (CET-4) and Band 6 (CET-6). College students are expected to take the CET-4 at the end of their second year in college. Students are eligible to take the CET-6 only after they have passed the CET-4. The CET-4 is usually required, whereas the CET-6 is optional. Both the CET-4 and the CET-6 are held twice a year: in June, at the end of the spring semester, and in December, at the end of the fall semester. Students can take the test multiple times while in college.

Traditionally, the CET consisted of five sections: listening, reading, vocabulary and structure, cloze (error-correction and/or question answering), and writing. The total possible score was 100, with a mean of 72 and a standard deviation of 12. Certificates were issued to examinees who achieved a score of over 60. However, with the issuance of the National College English Curriculum Requirements (NCECR, Ministry of Education, 2004), the CET underwent some major changes. For example, as of June 2005, a new scoring system was adopted, which has a maximum score of 710, a mean of 500, and a standard deviation of 70. Score reports are provided to examinees in order to provide feedback to guide their subsequent efforts to learn English. More importantly, as the NCECR aims to promote communicative English skills such as listening and speaking, the contents of the CET were adjusted accordingly in 2006. For example, the weight of the listening section was increased from 20% to 35%; the vocabulary and structure section, which originally constituted 15% of the total score, was removed. Skimming and scanning was introduced to the reading section, which now constitutes35% of the total score. Translation (or short-question answering) became a regular section and constitutes 5% of the total score, and the weight of the writing section remains the same at 15%.

The CET is regarded as the most influential English test in China, and it is the language test administered to the most students nationwide (Jin & Yang, 2006). According to Jin (2008, p. 2), the chair of the CET committee, "the CET is now taken by almost every college and university non-English-major student in China," and as many as 12 million students took the test in 2006. Before the reform of the CET in 2005, as reported by Yu (2005), 81.7% of Chinese universities regarded passing the

CET-4 as a precondition for earning a bachelor's degree. With the new scoring system introduced in 2005, the CET certificate is not issued any more; however, some universities still require a minimum CET-4 score, such as 426, in order for students to receive a bachelor's degree. Employers take a similar view: "CET certificates [or scores] have become a nationally recognized credential for employment of college and university graduates" (Jin, 2008, p. 4). For instance, CET scores have even been used to determine whether college students are eligible for residence permits in some major cities (Jin, 2008).Overall, the CET has had a far-reaching educational and social impact in China, and its impact has become a correspondingly important and highly contentious issue not only in academia but also in Chinese society (Jin, 2008; Yang, 2003). Many assert that the CET has driven the enforcement of the national curriculum and made a massive contribution to college English teaching in China (e.g., Gu, 2005; Yang, 2003). Meanwhile, the CET is criticized for inducing students to focus their English learning efforts on the test, i.e., to "study to the test," thus turning college English education into CET preparation (e.g., Cai, 2005; Chen, 2008; Han, Dai, & Yang, 2004).

Wall (1997) defined test impact as "any of the effects that a test may have on individuals, policies or practices, within the classroom, the school, the educational system or society as a whole" (p. 291). Among the many stakeholders, students are probably the most important group, as major decisions are made about them based on their test results (Kirkland, 1971). However, most of the studies on the impact of the CET (e.g., Chen, 2007; Gu, 2005; Hua, 2006; Wang, Wang, & Liu, 2005) have focused on teachers. Overall, there is a dearth of empirical evidence in regard to the impact of the CET on students; and, in particular, data is scarce in regard to their affective conditions, such as feelings, attitudes, and moods. Thus, the present study investigates students' perceptions of the impact of the CET on their English-learning practices and on their affective conditions, in order to enable school administrators, teachers, parents, test designers, and policy makers, as well as students, to become better informed about how the CET influences students. The current study focuses on the CET-4, as it is more influential than the CET-6.

## Literature Review

In the past several decades, the impact of tests has been the subject of considerable attention from educators and researchers—especially in the field of language testing. The term frequently used in language testing is "washback," defined by Hughes as"the effect of testing on teaching and learning" (1989, p. 1). In this paper, "washback" and "impact" are used interchangeably (Andrews, Fullilove, & Wong, 2002). The washback hypotheses proposed by Alderson and Wall (1993) provide clear guidelines on the areas that might be influenced by washback. For example, a test will influence teaching and learning; a test will influence what teachers teach and how they teach; a test will influence what learners learn and how they learn; and a test will influence the rate, sequence, degree, and depth of teaching and learning. Hughes (1993) identified three key mechanisms within the washback process, i.e., participants, process, and products. The participants are those whose perceptions of their work may be affected by a test, including students, teachers, administrators, materials developers, and publishers; the process is any action taken by the

participants that contributes to the learning process; and the products refer to what is learned and the quality of the educational outcomes. According to Hughes (1993), a test will first influence the participants' perceptions and attitudes, then how they perform, and finally the learning outcomes. Furthermore, washback can be considered positive (beneficial) or negative (harmful) (Taylor, 2005).Positive washback encourages good teaching and learning practices, whereas negative washback encourages bad teaching and learning practices.

Language testing researchers have conducted numerous studies examining the nature of washback, how it works, and its effects. Wall and Horak (2007), for example, offered a summary showing that a variety of studies have examined the impact of international tests such as the Test of English as a Foreign Language (TOEFL) (Alderson &Hamp-Lyons, 1996; Hamp-Lyons & Brown, 2005), the First Certificate in English (FCE) (Tsagari, 2006), the International English Language Testing System (IELTS) (Green, 2006; Hawkey, 2006; Hayes & Read, 2004), the Hong Kong Certificate of Education Examination in secondary schools (Cheng, 1997), the National Matriculation English Test (NMET) in China (Qi, 2007), the College Entrance Examination in English as a Second Language in Japan (Watanabe, 1996), and the Certificate in Spoken and Written English (CSWE) in the Australian Adult Migration Program (Burrows, 2004). The most common findings are these: tests do influence teaching and learning; tests tend to bring faster and more changes in teaching content than in teaching methods; and washback is more complicated than was first thought. However, as Wall (2000) observed, most studies have focused on how testing influences classroom teaching, where as studies on how testing influences students' learning and their behaviors are relatively few.

Of the studies on the CET washback, Gu's (2005) study is probably the most comprehensive one. Based on classroom observations at a local university and questionnaire surveys conducted nationwide in China, she found that most stakeholders perceived more positive washback of the CET on classroom teaching and learning. Specifically, the CET had great influence on teaching content, teaching pace, and teachers' attitudes towards teaching than on the teaching methods. Further, the CET was important in motivating schools to adhere to the national curriculum and induced school administrators to attach greater importance to English courses. However, some negative washback was also perceived, such as a faster teaching pace, the use of coaching materials in class, and being unable to complete the textbook materials (Gu, 2005).

To investigate the washback of the new CET, Gu, Yang, and Liu (2011) revisited the classrooms of the three college English teachers who participated in the classroom observation portion of Gu's 2005 study. They concluded that the essential mode of college English teaching, being teacher-dominated, remained the same over years, though the CET has undergone noticeable changes. However, the teachers did spend more time on listening, skimming and scanning, and translation, i.e., the skills that have more weight in the new CET than they had in the earlier version. This follow-up study agrees with the previous baseline study (Gu, 2005) that the CET has a greater influence on the content taught than on teaching methods.

Huang and Yang (2002) studied the washback of the CET in 11 universities in China. They found that the CET exerted an influence on various aspects of college

English teaching and learning, and that the majority of teachers and students believed the CET yielded more positive washback than negative. They also concluded that different types of institutions and students experienced different intensities of washback. For example, the higher-ranking universities perceived less influence compared to the lower-ranking universities, and the first-year students perceived less influence than the second-year students. Hua (2006) examined the washback of the CET in three teachers' colleges. She found that the CET did not exert much influence on how college English teachers taught when students were in their first three semesters of college. However, as the CET approached during the fourth semester, the teaching strategies, teaching materials, and teaching activities in the class all focused on preparing the students to pass the CET. In addition, Li (2009) examined the impact of the CET writing section on the teaching of writing at a university in China. She observed that the CET did not change the way teachers taught English writing, probably due to the relatively low requirement of the CET writing section, its restrictive testing format, teachers' lack of training on how to teach writing, and the large class size.

Overall, in the literature on CET washback, researchers have tended to focus on the impact of the CET on teaching activities, whereas students' perceptions of the CET have met with scant attention. Students are actually the primary stakeholders in testing situations, as it is the student "whose status in school and society is determined by test scores and the one whose self-image, motivation, and aspirations are influenced" (Kirkland, 1971, p. 307). Rea-Dickins (1997) also contended that students are perhaps the most important stakeholders and "their views are among the most difficult to make sense of and to use" (p. 306). Furthermore, most studies have focused on academic factors, whereas students' affective conditions have been neglected. It is, therefore, important to directly assess how students feel about the impact of the CET, both in terms of their English-learning practices and affective conditions.

In an extensive literature review, Kirkland (1971) concluded that tests could influence factors such as a student's self-concept, motivation, level of aspiration, study practices, and anxiety. First, test scores influence a student's self-concept. Depending on a student's opinion regarding the accuracy of the test results, opinion of his/her performance on the test, capabilities, and other individual characteristics, tests can have a positive or negative influence on a student's self-concept. Second, the stakes of a test, the frequency with which it is given, and expectations of success or failure on the test can influence a student's motivation in regard to it. Third, level of aspiration refers to the level of achievement that a student expects to reach, and it is also related to both self-concept and motivation. Successful performance on a test increases the level of aspiration but failure on a test decreases it. Fourth, study practices refer to the ways in which a student studies in preparation for a test. It has been found that different types of tests, such as open-book versus closed-book, multiple-choice versus essay questions, influence a student's study practices differently. Finally, anxiety and tension are always associated with taking tests. Those who anticipate encountering difficulties during the test may experience more anxiety than those who have no such expectations. In particular, Harlen and Deakin-Crick (2003) reviewed the impact of tests on student motivation, finding a complex

interaction between motivation and other factors, such as effort, goal orientation, locus of control, self-efficacy, sense of self as a learner, self-esteem, self-regulation, and interest.

Based on the literature, we propose that the CET impacts students in two ways: academic and affective. The academic impact refers to the CET's influence on students' English-learning behavior, as this pertains to learning content(i.e., what students study) and learning methods (i.e., how they study), whereas affective impact refers to the CET's influence on students' affective conditions, such as goal orientation, motivation, self-efficacy, and anxiety.

## Methods

### Context and Participants

This study took place at a university in Beijing, one of the high-ranking universities in China. As this university requires its students to have high college entrance examination scores in order to enroll, its students' English language proficiency is generally higher than the national average. All the students are required to take English courses consecutively for four semesters in their first two years. At the end of the fourth semester, students take the CET-4, and those who fail will retake the test during the rest of their college years. Those who pass the CET-4 usually will proceed to take the CET-6. However, this university does not require students to achieve a certain score on the CET-4 in order to receive a bachelor's degree.

In May 2008, a few weeks before the CET-4 was scheduled, 150 students completed a questionnaire that asked them how they felt about the impact of the CET-4. Of the students who provided demographic data, 109 were female and 38 were male; there were 49 first-year students, 56 second-year students, 41 third-year students, and 2 fourth-year students. At the time of the survey, the third-year and fourth-year students had already taken the CET-4, the second-year students were scheduled to take the CET-4 in a few weeks, and the first-year students would not take the CET-4 until a year later.

### Instrumentation and Data Collection

A questionnaire was constructed to solicit students' perceptions of the impact of the CET-4.(Unless otherwise specified, the CET refers solely to the CET-4 in the subsequent section and the questionnaire). First, we drafted items under the categories as set out in the previous literature review (e.g., Alderson & Wall, 1993; Gu, 2005; Huang & Yang, 2002; Kirkland, 1971), such as learning content, learning methods, goal orientation, motivation, self-efficacy, and anxiety. Second, two outside experts were invited to comment on whether the questionnaire statements examined students' perceptions of test impact as informed by the literature. For example, the experts suggested adding "as a result of taking or preparing for the CET" to the items about self-efficacy, so that participants would be able to appropriately attribute the change of their self-efficacy to their experience with the CET. The experts also suggested revisions in regard to improving the clarity, accuracy, and independence of the items. For example, the item "I am a better learner of English because of the CET" was criticized as being too general. As suggested by the experts,

this item was changed to "Taking or preparing for the CET influences the way I learn English."Finally, we translated the English questionnaire into Chinese. A Chinese–English bilingual researcher was invited to evaluate the translation, and minor revisions in regard to wording were made based on this researcher's suggestions.

Next, we ordered the items randomly for the final draft of the questionnaire. The first part of the questionnaire comprised demographic information, whereas the second part comprised 4-option, forced-choice Likert-type items asking students to select whether they strongly agreed, agreed, disagreed, or strongly disagreed with each statement about the impact of the CET. Copies of the questionnaire, now rendered in Chinese, were distributed to 150 undergraduate students with the help of their English instructors. All the 150 students completed and returned the questionnaires to the researchers.

## Results and Discussion

### Learning Content and Learning Methods

Five items related to students' perceptions of how the CET impacted their English-study behavior. Table 1 shows the percentage of the students who strongly agreed (SA), agreed (A), disagreed (D), or strongly disagreed (SD) with each item. The items are listed in descending order of the overall percentage of SA and A responses. In other words, the item listed at the top has the highest percentage of SA and A responses, whereas the item at the bottom has the lowest percentage of SA and A responses. This organizational principle applies to all the tables in this paper.

Table 1
*Learning Content*

| Items | SA | A | D | SD |
|---|---|---|---|---|
| I will work hard to practice English speaking if it is required on the CET. | 21.3 | 56.7 | 20 | 2 |
| I will spend more time practicing English listening if listening gets heavier weight in the CET. | 21.3 | 56.7 | 17.3 | 4.7 |
| I am more attentive in the class if the teacher lectures on contents related to the CET. | 12.7 | 63.3 | 20.7 | 3.3 |
| I pay more attention to the content that is related to the CET. | 9.3 | 54 | 32 | 4.7 |
| I pay more attention to the words that are labeled as CET vocabulary. | 8.7 | 45.6 | 38.3 | 7.4 |

**Note: SA =Strongly Agree, A = Agree, D = Disagree, SD = Strongly Disagree. All the numbers in the cells are percentages.**

It can be seen that 78% of the students agreed or strongly agreed that they would work hard to practice English speaking if speaking were required on the CET. Speaking was added to the CET in 1999 as an optional section and has remained so. Only when students have achieved an advanced-level rating of 550 for CET-4 or 520

for CET-6, do they become eligible to take the CET speaking test (Zhang & Elder, 2009). According to the students' responses, if speaking were to become a required section of the CET, the majority of students would place more emphasis on developing their English-speaking ability. Likewise, 78% of students agreed or strongly agreed that they would spend more time practicing English listening if it were more heavily weighted in the CET. In 2006, the weight given to the listening section was raised from 20% to 35% with the purpose of encouraging students to develop communicative language skills. Though causal relations are difficult to make here, it is reasonable to assume that the changes in the CET with regard to listening skills affected students' attitudes regarding improving their communicative skills in English.

Similarly, 76% of the students agreed or strongly agreed that they would be more attentive in class if the teacher were to lecture on content related to the CET. In a related CET washback study, Li (2009) reported that students sometimes requested teachers to coach them on the CET in class, especially when the test was approaching. Therefore, some English teachers complained that the CET interfered with their regular classroom teaching (Gu, 2005; Hua, 2006). Likewise, 63% of students agreed or strongly agreed that they would pay more attention to the content that is related to the CET, which again tends to confirm the CET's influence on learning content. Furthermore, more than half the students agreed or strongly agreed that they would pay more attention to the words labeled as CET vocabulary. The CET Committee publishes manuals with words that might appear in their tests; thus, in the college English textbooks and the CET coaching materials, some words are labeled as CET-4 vocabulary and some more challenging ones are labeled as CET-6 vocabulary. Therefore, it is likely that students would pay particular attention to the words labeled as CET vocabulary.

Table 2 summarizes the four items related to learning methods. Over 75% of the students agreed or strongly agreed that they had taken or would take the CET coaching classes, and over 60% would buy or had bought the CET coaching materials. After-school CET coaching classes have become pervasive in China. Such classes not only provide intensive training on how to learn English but more importantly they offer training on test-taking strategies. Furthermore, a large number of CET coaching materials are on the market for students to purchase. However, the quality of the CET coaching classes and materials vary greatly, and they also add considerably to the financial burden of college students. Still, the results of this study show that more than half of the participating students would resort to CET coaching classes and/or purchase CET coaching materials.

Around 45% of the students agreed or strongly agreed that they would try any learning method that might help them perform better on the CET. Still, only about 29% of the students thought that taking or preparing for the CET influenced the way they learned English. There is consistent evidence showing that tests have a stronger influence on teaching content than on teaching methods (e.g., Cheng, 1997; Gu, 2005; Li, 2009). Accordingly, the current study shows that the CET tended to change the content the students studied more than the ways in which they studied it. Traditionally, English learning in China tends to be test-oriented, book-centered, with plenty of drills and exercises that emphasize rote memorization rather than

communicative skills (Rao, 2001). This particular learning style is the outcome of the sociocultural, economic, and historical setting of China (Li & Su, 2006), and thus it is likely to be difficult to change in a short time.

Table 2
*Learning Methods*

| Items | SA | A | D | SD |
|---|---|---|---|---|
| I have taken or will take the CET coaching classes. | 24.7 | 50.7 | 17 | 7.3 |
| I will buy or have bought CET coaching materials. | 13.3 | 48.7 | 31 | 7.3 |
| I would like to try any learning methods that can help me perform better on the CET. | 4.7 | 38.9 | 48 | 8.1 |
| Taking or preparing for the CET influences the way I learn English. | 6 | 22 | 43 | 29.3 |

Note: SA =Strongly Agree, A = Agree, D = Disagree, SD = Strongly Disagree. All the numbers in the cells are percentages.

**Goal Orientation and Motivation**

Goals provide students with direction and a purpose for engaging in an activity (Pintrich&Schunk, 1996), and goal orientations constitute students' reasons for engaging in academic tasks (Anderman, Austin, & Johnson, 2002). Three items are related to goal orientation. As shown in Table 3, over half of the students agreed or strongly agreed that taking or preparing for the CET led them to have clearer English-learning goals. However, only one third of them agreed or strongly agreed that passing the CET was their major driving force for learning English, and only 20% agreed or strongly agreed that passing the CET was their major purpose for learning English.

Table 3
*Goal Orientation*

| Items | SA | A | D | SD |
|---|---|---|---|---|
| Taking or preparing for the CET makes me have clearer goals in learning English. | 10 | 45.3 | 38 | 6.7 |
| To pass the CET is my major driving force in learning English. | 5.3 | 26.7 | 40.7 | 27.3 |
| To pass the CET is my major purpose for learning English. | 5.3 | 14.7 | 51.3 | 28.7 |

**Note: SA =Strongly Agree, A = Agree, D = Disagree, SD = Strongly Disagree. All the numbers in the cells are percentages.**

The historical civil service exam in China has dominated the history of the Chinese educational system, and it still influences schooling practices today (Suen & Yu, 2006). Specifically, English-language tests are used as gate-keeping devices for access to general employment and higher education in China (Ross, 2008). It is, therefore, not surprising that English-language education is intensively test-oriented

in China. Many students learn English for the sake of taking the tests instead of for using the language for real purposes. Therefore, for many students, the CET appears to be one of their goals for learning English in college. In addition to the CET, other English tests that college students can take include the graduate school entrance examination in China and international English tests such as the Test of English as a Foreign Language (TOEFL), the Test of English for International Communication (TOEIC), the International English Language Testing System (IELTS), and the Business English Certificate (BEC) (Jin, 2008). According to Hua (2006), 70% of the students reported that they were motivated to learn English by the CET; 14% reported that their motivation was to pass the graduate school entrance examination, and 10% were motivated to improve their English language abilities. In Hua's view, these results reflect the fact that the students in her study were likely to remain in China as public school teachers, such that few would have a need to present scores from international English tests. In the current study, however, the participating students were from a high-ranking university, and many of them intended to enroll in graduate schools in China or abroad. This partially explains why most of them did not regard passing the CET as their major purpose for learning English.

A concept related to goal orientation is motivation, which "is the process whereby goal-directed activity is instigated and sustained" (Pintrich & Schunk, 1996, p. 4). Motivation is a broader concept, and different theoretical models have been developed to describe it. Motivation in the present study is mainly operationalized by how much effort the students were willing to make to pass the CET. Table 4 shows the results of 12 items related to motivation. To begin with, over 80% of the students agreed or strongly agreed that taking or preparing for the CET made them more motivated to learn English. Seventy-four percent agreed or strongly agreed that they spent more time learning English because of taking or preparing for the CET. More than half of the students agreed or strongly agreed that the CET made them feel that English was a very useful tool, and taking or preparing for the CET made them feel that English learning was more important.

Table 4
*Motivation*

| Items | SA | A | D | SD |
|---|---|---|---|---|
| Taking or preparing for the CET makes me more motivated to learn English. | 17.3 | 66 | 14 | 2.7 |
| I spend more time learning English because of taking or preparing for the CET. | 22 | 52 | 20.7 | 5.3 |
| In order to prepare for the CET, I spend more time memorizing English words. | 14 | 58 | 24 | 4 |
| In order to prepare for the CET, I spend more time watching English movies. | 12 | 58.7 | 26 | 3.3 |
| In order to prepare for the CET, I spend more time listening to English broadcasts. | 9.3 | 60 | 28 | 2.7 |
| In order to prepare for the CET, I spend more time practicing English–Chinese translation. | 8.7 | 58.7 | 30.7 | 2 |

| | | | | |
|---|---|---|---|---|
| In order to prepare for the CET, I spend more time practicing English writing. | 10 | 54.7 | 30 | 5.3 |
| In order to prepare for the CET, I spend more time reading English newspapers. | 7.3 | 51.3 | 33.3 | 8 |
| Taking or preparing for the CET makes me pay more attention to English use in real life. | 13.3 | 39.3 | 36.7 | 10.7 |
| The CET makes me feel that the English language is a very useful tool. | 11.3 | 41.3 | 36 | 11.3 |
| Taking or preparing for the CET makes me feel that learning English is more important. | 8.1 | 43.6 | 40.9 | 7.4 |
| In order to prepare for the CET, I spend more time learning English and American literature. | 3.3 | 24 | 52.7 | 20 |

Note: SA =Strongly Agree, A = Agree, D = Disagree, SD = Strongly Disagree. All the numbers in the cells are percentages.

However, the extra time students stated they would spend on specific language skills varied. For example, as many as 72% of students agreed or strongly agreed that they would spend more time memorizing English words, which provides evidence that rote memorization is deeply rooted in Chinese students' English-learning practices (Li, 2005). In addition, approximately 70% of the students agreed or strongly agreed that they would spend more time watching English movies and listening to English broadcasts. This agrees with Hua's (2006) observation that students spent more time on listening and speaking than on other skills, possibly as a result of the greater focus the new version of the CET on communicative skills. Less than 60% of the students agreed or strongly agreed that they would spend more time reading English newspapers. It seems that students learned English not only from books but also from other channels; however, newspapers were not as popular as movies or broadcasts for college students' English learning. Those who agreed or strongly agreed that they would spend more time practicing English-Chinese translation totaled 66.4%, whereas 64.7% agreed or strongly agreed that they would spend more time practicing English writing. Translation and writing are assessed in the CET but each has a relatively low weight, with the former accounting for 5% and the latter accounting for 15% of the total score. About 52.6% of the students agreed or strongly agreed that they would pay more attention to English use in real life. This relatively low percentage might be explained by the fact that it is difficult for standardized language tests to directly assess authentic language use in real life, especially when most of the questions are framed in a multiple-choice format. Finally, only 27.3% of the students agreed or strongly agreed that they would spend more time learning English and American literature. One potential reason for this low percentage could be that English literature is not directly tested in the CET.

**Self-Efficacy and Anxiety**

Self-efficacy is defined as how people feel about their ability to produce designated levels of performance that exercise influence over events that affect their lives (Bandura,1994). As shown in Table 5, about half the students agreed or strongly

agreed that they felt more confident about their English reading and listening ability as a result of taking or preparing for the CET, whereas less than 40% felt the same way about their speaking and writing ability. Reading has always been a focus of the CET, whereas the weight given to listening was raised from 20% to 35% in 2006. It is plausible to expect, therefore, that students may have put more effort into reading and listening and consequently felt more confident about their reading and listening ability. In contrast, speaking is optional in the CET, and writing only accounts for 15% of the CET total score. It is reasonable to surmise, therefore, that this partially explains the relatively less increase of self-efficacy in terms of speaking and writing. Furthermore, around 47% of students felt more confident about their overall English proficiency, and about 42% felt more capable of using English in real situations as a result of taking or preparing for the CET.

Table 5
*Self-Efficacy*

| Items | SA | A | D | SD |
|---|---|---|---|---|
| I feel more confident about my English-reading ability as a result of taking or preparing for the CET. | 3.3 | 50 | 40 | 6.7 |
| I feel more confident about my English-listening ability as a result of taking or preparing for the CET. | 7.3 | 42.7 | 42 | 8 |
| I feel more confident about my overall English proficiency as a result of taking or preparing for the CET. | 4.7 | 42.7 | 46 | 6.7 |
| I feel more able to use English in real situations as a result of taking or preparing for the CET. | 4 | 38 | 48 | 10 |
| I feel more confident about my English-speaking ability as a result of preparing for the CET. | 4.7 | 32.7 | 48 | 14.7 |
| I feel more confident about my English-writing ability as a result of taking or preparing for the CET. | 2 | 34 | 54 | 9.5 |

Note: SA =Strongly Agree, A = Agree, D = Disagree, SD = Strongly Disagree. All the numbers in the cells are percentages.

Anxiety, defined as the "subjective feeling of tension, apprehension, nervousness, and worry associated with an arousal of the automatic nervous system" (Spielberger, 1983, p. 1), is also an important aspect of the test's impact. As shown in Table 6, over 80% of the students agreed or strongly agreed that they were under greater pressure because of taking or preparing for the CET, and almost 70% agreed or strongly agreed that they were experiencing more anxiety in terms of learning English. A certain level of anxiety may promote learning, though too much may function as an obstacle in this regard (Kirkland, 1971). Thirty-eight percent of the students agreed or strongly agreed that taking or preparing for the CET made them feel that they had failed in their efforts to learn English. Around one third of the students agreed or strongly agreed that they felt more frustrated with learning English and had become more afraid of learning English. This is somewhat in contrast with the increasing self-efficacy they reported in terms of their English

ability, as illustrated in Table 5. This contrast may represent the actual situation regarding students' perceptions of the impact of the CET. On the one hand, students felt more confident about their English ability as a result of putting more effort into preparing for the CET. On the other hand, the test preparation put them under great pressure, such that they became more anxious about and frustrated with learning English.

Table 6
*Anxiety*

| Items | SA | A | D | SD |
|---|---|---|---|---|
| I am under greater pressure to learn English because of taking or preparing for the CET. | 18.4 | 65.3 | 14 | 2 |
| Taking or preparing for the CET makes me feel more anxious about learning English. | 10.7 | 58 | 28 | 3.3 |
| Taking or preparing for the CET makes me feel that I have failed in my efforts to learn English. | 6.7 | 31.3 | 53 | 9.3 |
| Taking or preparing for the CET makes me more frustrated with learning English. | 6 | 27.3 | 53 | 13.3 |
| I am more afraid of learning English because of taking or preparing for the CET. | 6.1 | 25 | 56 | 12.8 |

Note: SA =Strongly Agree, A = Agree, D = Disagree, SD = Strongly Disagree. All the numbers in the cells are percentages.

## Conclusion

Based on evidence from the questionnaire survey, the CET seems to have had a pervasive impact on the participating students in this study. First, the CET has a greater impact on learning content than on learning methods. The CET seems to be effective in directing students' attention to what they need to learn in order to pass it. This finding is consistent with previous literature showing that tests have a greater impact on the content taught than on the methods used to teach it. Second, over half of the students felt that the CET had caused them to clarify their English-learning goals, and over 80% of the students were more motivated to make a greater effort to learn English. A pattern seemed to emerge whereby the students usually expressed their willingness to match their effort to learn respective language skills in accordance with their weight in the CET. Therefore, many students were willing to put more effort into listening and reading and less effort into writing and speaking. Third, about half the students reported a higher level of self-efficacy in regard to their overall English ability and to different English skills as a result of preparing for or taking the CET. Despite the increasing self-efficacy, however, almost 70% of the students reported that they felt more pressure and anxiety as a result of preparing for or taking the CET, and around one third of the students felt more frustrated with learning English and became more afraid of learning English. Overall, the findings of this study provide some evidence for the intensity and range of the impact of the CET on students' English learning, both in terms of their English-learning practices and affective conditions.

The English education in China has been characterized by "concentration on intensive learning, preoccupation with examinations of grammatical structures, memorization and rote learning of vocabulary, and lack of attention to more communicative skills" (Harvey, 1990, cited in Qi, 2005, p. 145). And, in fact, a frequent criticism of the CET is that it promotes such practices as rote memorization. As with many other large-scale standardized tests, for the purpose of efficiency, the CET relies heavily on multiple-choice items. As a result, some lower-level language skills such as vocabulary and grammar are emphasized, whereas higher-level communicative skills may not receive adequate attention. Given the importance of the CET in college English education, it is not surprising to find that many undergraduate students focus on reciting English words and practicing English grammar exercises (Wang, 2010). Fortunately, the CET committee has taken some steps to address this issue (Zheng & Cheng, 2008). For instance, since 2006, the test has given more weight to communicative skills and removed the direct assessment of vocabulary and sentence structure from the multiple-choice items. The current study has found that students seem to be sensitive to what is assessed in the CET and the weight it gives to different language skills. Given the CET's powerful impact on college English education, it is important that the CET committee keep reforming the test and adopt more authentic measures of English-language ability in order to encourage students to take more interest in English-language use in real-world contexts.

As described by Shohamy (2001, p. 113), high-stakes tests can have detrimental effects on individuals "as they can create winners and losers, successes and failures, rejections and acceptances." In the present study, most of the students were motivated to learn English as a result of preparing for or taking the CET, yet the CET also made many of them feel more anxious about and frustrated with their efforts to learn English. These negative emotional effects are likely to be related to the high stakes inhering in success or failure on the CET. Although this was not intended by the CET committee, some Chinese universities use the CET results as a basis for awarding a bachelor's degree and some employers also require the CET score reports. The debate over whether test developers are responsible for how tests are used has long been in progress, and Jin (2008, p.1) used the phrase "powerful tests, powerless test designers" to describe this difficult situation. As recommended by the Code of Ethics for the International Language Testing Association (2000), "language testers shall regularly consider the potential effects, both short and long term on all stakeholders of their projects, reserving the right to withhold their professional services on the grounds of conscience." Although the CET committee cannot control how the test results are used by universities or employers, more empirical studies should be conducted to validate or invalidate certain ways of using the CET test results, and the CET committee would be the best agent to coordinate or promote such efforts.

**Limitations and Future Studies**

The university where the current study took place had a relatively relaxed policy in regard to the CET. For example, if undergraduates are not able to achieve a certain score on the CET-4 before graduation, they can still earn their bachelor's degrees, as

long as they pass a school-based English proficiency test that is less challenging than the CET-4.However, for the universities that use earning a certain score on the CET as a degree requirement, the CET is likely to have higher stakes for their students and thus may bring stronger effects. The stakes of the CET are thus context-based. In addition to the premise that the perceptions of the CET could be different at different universities, the impact of the CET on students may also interact with students' individual characteristics, such as their English proficiency, gender, social economic status (SES), and affective conditions. It is expected that students with lower English proficiency may experience a correspondingly greater negative impact from the CET (Green, 2006). However, given that some of the participating students had not yet taken the CET-4, there is a lack of a common measure of students' English proficiency in the present study. Also, as the present study collected data from only one university with students having similar college entrance examination scores, there is little variation in students' English proficiency. Therefore, it is necessary to expand this research to involve students from different universities with heterogeneous English-proficiency levels and family backgrounds in order to explore individual differences related to their perceptions.

To understand the general perceptions of the impact of the CET, we recruited participants from different grade levels at college. The third-year and fourth-year students had taken the CET, whereas the first-year and the second-year students were preparing to take the CET when they responded to the survey. Many of the survey items were thus designed to cater to the wide range of participants' experience with the CET, such as "I have taken or will take the CET coaching classes." If a study goal is to compare the "veteran" CET-takers' versus "future" CET-takers' perceptions, it would be wise to build different versions of items for these two groups of participants. For example, the item could be written as "I have taken the CET coaching classes" for "veteran" CET-takers and "I will take the CET coaching classes" for "future" CET-takers. Another limitation is that unequal sample sizes by gender and school year were used in this study. More male students and more fourth-year students should be included in further studies in order to gain a more balanced perspective on the CET.

Finally, students' self-reported responses to the survey are the only evidence used in this study. It would be helpful to collect other types of evidence, such as observations and interviews, to gain a more in-depth understanding of how and why students had particular perceptions of the impact of the CET. Nevertheless, the current study constitutes only a first step in investigating the impact of the CET on students at one university. Additional long-term and follow-up studies at different types of universities are needed in order to learn about the nature of the CET's impact in a variety of settings.

**References**

Alderson, J. C., &Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing, 13*(3), 280–297.

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115–129.

Anderman, E. M., Austin, C. C., & Johnson, D. M. (2002). The development of goal orientation. In A. Wigfield, & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 197–220). San Diego, CA: Academic Press.

Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback: A case study. *System, 30*(2), 207–223.

Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 71–81). New York:, NY Academic Press.

Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian Adult Migration Program. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing – Research contexts and methods* (pp. 113–128). Mahwah, NJ: Lawrence Erlbaum.

Cai, J. (2005). *ELT at the tertiary level in China: Review, reflection and research.* Shanghai, China: Fudan University Press.

Chen, F. (2007).The washback effect of College English Test Band 4 on curricular planning and instruction. *CELEA Journal, 30*(1), 19–29.

Chen, H. (2008). The influence of the College English Test Band 4 and Band 6 on College English teaching. *Chinese and Foreign Education Research, 10*, 35–37.

Cheng, L. (1997). Does washback influence teaching? Implications for Hong Kong. *Language and Education, 11*(1), 38–54.

Green, A. (2006). Watching for washback: Observing the influence of the International English Language Testing System academic writing test in the classroom. *Language Assessment Quarterly, 3*(4), 333–368.

Gu, X. (2005). *Positive or negative? An empirical study of CET washback on college English teaching and learning in China* (Unpublished doctoral dissertation). Shanghai Jiaotong University, Shanghai, China.

Gu, X., Yang, Z., Liu, X. (2011). *A longitudinal case study of CET washback on college English classroom teaching in China*. Paper presented at the 33rd Language Testing Research Colloquium, Ann Arbor, MI.

Hamp-Lyons, L., & Brown, A. (2005). *TOEFL impact study report* (Unpublished report). University of Melbourne, Victoria, Australia.

Han, B., Dai, M., & Yang L. (2004). Analyzing the problems of the College English Test based on a survey. *Foreign Languages and Their Teaching, 179*(2), 17–23.

Harlen, W., & Deakin-Crick, R. (2003).Testing and motivation for learning. *Assessment in Education, 10*(2), 169–208.

Harvey, P. (1990). A lesson to be learned: Chinese approach to language learning. In Dzau, Y. F. (Ed.), *English in China* (pp.168–174). Hong Kong: API Press.

Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progettolingue 2000.* Cambridge: Cambridge University Press.

Hayes, B., & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. In L. Cheng, Y. Watanabe, & A.

Curtis (Eds.), *Washback in language testing – Research contexts and methods* (pp. 97–111). Mahwah, NJ: Lawrence Erlbaum.

Hua, S. (2006).An empirical study of washback from CET-4 on college English teaching and learning. *CELEA Journal, 29*(1), 54–59.

Huang, D., & Yang, L. (2002). Analyzing the problems of the College English Test based on a survey. *Foreign Language Testing and Research, 4*, 288–293.

Hughes, A. (1989). *Testing for language teachers*. Cambridge, UK: Cambridge University Press.

Hughes, A. (1993). *Backwash and TOEFL* (Unpublished manuscript). University of Reading, England.

International Language Testing Association. (2000). *Code of ethics*. Retrieved June 1, 2011 from http://www.iltaonline.com/images/pdfs/ILTA_Code.pdf

Jin, Y. (2008). Powerful tests, powerless test designers? Challenges facing the college English test. *CELEA Journal, 31*(5), 1–11.

Jin, Y., & Yang, H. (2006). The English proficiency of college and university students in China: As reflected in the CET. *Language Culture and Curriculum, 19*(1), 21–36.

Kirkland, M. C. (1971). The effect of tests on students and schools. *Review of Educational Research, 41*(4), 303–350.

Li, H. (2009). Are teachers teaching to the test? A case study of the College English Test (CET) in China. *International Journal of Pedagogies and Learning, 5*(1), 25–36.

Li, X. (2005). An analysis of Chinese EFL learners' beliefs about the role of rote learning in vocabulary learning strategies. *Asian EFL Journal, 7*(4). Retrieved from http://www.asian-efl-journal.com/xiuping_11-05_thesis.pdf

Li, Z., & Su, W. (2006). Socio-cultural background and Chinese English majors learning styles. *CELEA Journal, 29*(3), 72–81.

Ministry of Education, P. R. China. (2004). *National college English curriculum requirements.*

Pintrich, P., & Schunk, D. (1996). *Motivation in education: Theory, research & applications*. Englewood Cliffs, NJ: Prentice-Hall.

Qi, L. (2005). Stakeholders' conflicting aims undermine the washback functions of a high-stakes test. *Language Testing, 22*(2), 142–173.

Qi, L. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. *Assessment in Education, 14*(1), 51–74.

Rao, Z. (2001). Matching teaching styles with learning styles in East Asian contexts.*The Internet TESL Journal*, *7*(7). Retrieved from http://iteslj.org/Techniques/Zhenhui-TeachingStyles.html

Rea-Dickins, P. (1997). So, why do we need relationships with stakeholders in language testing? A view from the UK. *Language Testing, 14*(3), 304–314.

Ross, S. J. (2008). Language testing in Asia: Evolution, innovation, and policy challenges. *Language Testing, 25*(1), 5–13.

Shohamy, E. (2001). The social responsibility of the language testers. In R. L. Cooper, E. Shohamy, & J. Walter (Eds.), *New perspectives and issues in educational language policy: A festschrift for Bernard DovSpolsky,* (Vol. in honor of Bernard Spolsky, pp. 113–130). Philadelphia, PA: John Benjamins North America.

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

Suen, H. K., & Yu, L. (2006). Chronic consequences of high-stakes testing? Lessons from the Chinese civil service exam. *Comparative Education Review, 58*(1), 46–65.

Taylor, L. (2005). Key concepts in ELT: Washback and impact. *ELT Journal, 59*(2), 154–155.

Tsagari, C. (2006). *Investigating the washback effect of the first certificate in English examination in Greek private language schools* (Unpublished doctoral dissertation). Lancaster University, Lancaster, UK.

Wall, D. (1997). Impact and washback in language testing. In C. Clapham, & D. Corson (Eds.), *Encyclopedia of language and education. Volume 7: Language Testing and Assessment,* (pp. 291–302). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Wall, D. (2000). The impact of high stakes testing on teaching and learning: Can this be predicted or controlled? *System, 28*(4), 499–509.

Wall, D., & Horak, T. (2007). Using baseline studies in the investigation of test impact. *Assessment in Education, 14*(1), 99–116.

Wang, W. (2010). Evaluation on washback effects of College English Test Band 4. *Sino-US English Teaching, 7*(2), 35–40.

Wang, X., Wang, J., & Liu, F. (2005). CET validity: An analytical study based on the feedback from the questionnaires. *Journal of Heibei Institute of Architectural Science and Technology (Social Science Edition), 22*(1), 109–110.

Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing, 13*(3), 318–333.

Yang, H. (2003). The 15 years of the CET and its impact on teaching. *Journal of Foreign Languages, 145*(3), 21–29.

Yu, W. (2005). The student survey and analysis on the reform of College English Test Band 4 and Band 6. *China College Teaching, 7*, 43–47.

Zhang, Y., & Elder, C. (2009). Measuring the speaking proficiency of advanced EFL learners in China: The CET-SET solution. *Language Assessment Quarterly, 6*(4),298–314.

Zheng, Y., & Cheng, L. (2008). Test review: College English Test (CET) in China. *Language Testing, 25*(3), 408–417.