The Asian ESP Journal

Autumn Edition

Volume 6 Issue 2

October 2010

Editors:

Dr. Paul Robertson

Prof. Roger Nunn

Prof. Winnie Cheng

Published by Time Taylor Press

Asian ESP Journal
A Division of Time Taylor International Ltd
Time Taylor International
Australia

http://www.asian-esp-journal.com

asianefl@gmail.com

Publisher: Dr. Paul Robertson
Chief Advisor: Dr. Roger Nunn
Chief editor. Prof. Winnie Cheng

ISSN 2206-0979 (Online)

Table of Contents

The Asian ESP Journal

Foreword

Prof. Winnie Cheng, Chief Editor

Welcome to the Autumn 2010 issue of The Asian ESP Journal! It is our pleasure to publish six selected papers reporting on the results of a range ESP studies from authors from Iran, China and Taiwan. The studies have examined a wide range of topics. Genres studied include research articles and textbooks for Medical Sciences students. ESP contexts studied include university entrance examinations and web-based instructional environment.

The first paper 'The status of theme in applied linguistic articles' by A.R. Jalilifar is a corpus-based study that compares the themes in 16 articles published in a local ELT journal and 16 in an international ELT journal. While similarities are found the in the theme types and thematic progression patterns, significant differences are found in the number and the context of use of different patterns of thematic progression in two sections of the articles: Introduction and Results and Discussion.

Chun-Chun Yeh's 'Citation practices in TESL research articles: A comparative study' is also concerned with research writing and corpus-based. It compares citation practices in TESL research articles published in a Taiwanese journal and an international journal, and finds both similarities and differences in the frequencies of citations, the types of citations used, and where in the articles that they tend to be used.

The third paper 'Authenticity in ESP/EAP selection tests administered at Iranian universities' by Gholam Reaz Kiani, Reza Ghafar Samar, Ramin Akbari, and Goodarz Alibakhshi is interested in finding out the extent to which Iranian university entrance examinations for MA/MSC and PhD students are authentic in terms of target language use situation tasks and learners' language needs. The findings of their survey research have important implications for the design of similar tests in other parts of the world.

Also from Iran, 'The impact of personality traits on the writing performance of Iranian EFL learners' by Mohammad Mehdi Soleimani and Mahmoud Reza Daryabari describes the relation between the test-taker's personality type and degree of success at accomplishing writing tasks. The findings, which show that extroverted language learners outperform introverted learners, are different from those of a previous study.

'Evaluation of SAMT ESP textbooks for the students of medical sciences' is a contribution from Iran by S.A. Razmjoo and Reza Raissi. Questionnaire surveys for students and instructors were administered to find out their respective views on the Medical Sciences textbooks produced by The Center for Studying and Compiling University Books in Humanities (SMAT) based on six criteria. The study shows no significant difference between students and instructors, who view the textbooks rather negatively in many of the criteria.

The sixth paper is by Zhu Ming entitled 'A case study of technology integration in an EFL classroom'. Using survey methods, the author finds that students in an EFL classroom are in favour of web-based instructional environment, and perceive benefits in developing their writing skills, communication skills, and learner autonomy, and in becoming self-

dependent responsible learners. The positive results will boost the adoption of web-based learning in China, where this mode of instruction is still quite new.

We hope you will enjoy reading the six papers and find the wide-ranging research findings and pedagogical implications useful for your academic and research pursuits!

Last but not least, we are grateful to a number of fellow colleagues for the generous contribution of their valuable time and effort in the review and proofreading of the six papers. They are the Associate Editors and Editors of the Asian ESP Journal. The proofreaders are:

1. Dr. Stacia Levy, University of the Pacific, Stockton, California, USA
2. Dr. Julia Penelope, USA, Unaffiliated
3. Dr. Cuijun Hu, Burnaby, Greater Vancouver, Canada
4. Mae Hurley, Language Analyst, Plain English Foundation, Sydney, Australia
5. Peyman Ghassemi Pour Sabet, Curtin University
6. Prof. Samir Diouny, Professor of Linguistics, Département de Langue et Littérature anglaises, Faculté des Lettres et des Sciences Humaines, Université Chouaib Doukkali
7. Norman Fewell, Associate Professor, Foreign Language Center, University of the Ryukyus, Okinawa, Japan

The Asian ESP Journal

The Status of Theme in Applied Linguistics Articles

A. R. Jalilifar

Shahid Chamran University of Ahvaz

Abstract

The ultimate goal of local writers of ELT articles may be successful publication of their articles in international journals. But their articles may be rejected due to problems with thematic organization which has a crucial role in readability and coherence of the articles. The present study aimed at comparing different thematic types and thematic progression patterns used in different rhetorical sections of international and local journals. Following Hallidayan (1985) categorization of theme and the revised model of thematic progression patterns proposed by McCabe (1999), this study incorporated a corpus of 32 experimental articles (16 from ELT and 16 from Roshd) in order to find out the possible similarities and differences in terms of thematic selection. The analysis showed overall similarities in both journals regarding different types of theme and patterns of thematic progression. These similarities can be attributed to the shared genre. But there were significant differences in the number and the context of the usage of different patterns of thematic progression in the introduction and results and discussion sections of both journals, indicating the descriptive nature of Roshd and argumentative nature of ELT journals. The results confirmed the need for informing local writers of English of the crucial role of thematic organization in the writing of ELT articles.

Keywords: Theme types; Thematic progression patterns; ELT journals; T-unit

1. Introduction

Theme is one of the textual choices that language producers make at the clause level (Halliday, 1994; Paltridge, 2006). It has a crucial role in the readability and in the appearance of a text and, in general, plays a central role in our every day communication. It sets up an environment in which the message can be interpreted and understood clearly (Matthiessen & Halliday, 1997, p. 19). It is both cognitive in that it refers to the world of experience and communicative in that it has a discoursal role (McCabe, 1999, p. 65) and correlates with the exposition of text (Fries, 1983).

There is a considerable agreement among linguists that theme acts as a cohesive device in the texts. According to McCabe (1999), thematic progression (TP), which refers to the relation of a theme to other themes and rhemes of the texts, provides continuing discourse and organizes cohesive text (p. 169). It constitutes a domain which helps the texts to be cohesive with respect to themselves and the contexts of situation.

Moreover, theme helps the speaker/writer to specify the place in the listener/ reader network of meaning where the meaning is to be incorporated as relevant (Matthiessen & Halliday, 1997, p.19). Fries (1995) believes that the kinds of meaning in the thematic position vary depending on the purpose of the writer/ speaker, who can manipulate the reactions of reader/ listener by changing the content of themes (p. 7).

Writers may lose effectiveness in their arguments due to problems with thematic selections, thematic progression, or both. And it is this problem that non-native writers of English, especially those who wish to publish in scholarly journals, may often grapple with. Many of these writers may suffer the feeling of exclusion and marginalization mainly because

their articles are rejected by international journals. So writing with an awareness of the thematic choices may help non-native writers of English to construct thematically appropriate texts and eventually lead to the ultimate goal of successful publication.

2. Literature Review

Despite its varying realizations from one language to another, theme is important in many of the world's languages, and a theme within the clause is a universal feature (McCarthy, 1991, p. 52). The importance of thematic choices in academic and non-academic texts has attracted the attention of many researchers in different languages (e.g., Baltazani, 2006; Ghadessy, 1995; Ji, 2002; Martinez, 2003; McCabe & Belmonte, 2001). Researchers have also investigated the theme/ rheme structure and its organization in writing in order to find out similarities and differences in the way thematic structures are employed by native and non-native writers of English. Some studies have analyzed the relevance of clause initial position in the organization of news reports (Gomez, 1994); others have looked into students' written essays from different disciplinary backgrounds (North, 2005; Wang, 2007). A more recent study investigated variations involved in the thematic choices when academic texts are translated (Jalilifar, 2009).

In studying academic articles, Whittaker (1995) analyzed eight academic articles, four on economics and four on linguistics, to describe thematic choices by authors. Results revealed that academic writing depends more heavily on relational processes and less on interpersonal themes due to the impersonal tone of academic articles. Belmonte and McCabe (1998) compared professional texts with the compositions of ESL students to investigate the significance of theme and rheme in the comprehension and production of texts. As expected, according to McCabe (1999), professional writers made use of different thematic patterns (linear, constant, split theme, and split rheme) to develop the topic. Ghadessy (1999) explored the thematic organization of 150 academic article abstracts. The study suggested that a

common method of development or point of departure is established through theme selection by professionals. The noticeable feature of all abstracts was the high frequency of simple and marked themes; in addition, the circumstances occurred more frequently than processes. Martinez (2003) compared the thematic realizations in the method and discussion sections of biology research articles, assuming that the differences between these sections manifest themselves in different thematic structures. She argued that in a specific genre, the linguistic choices will differ across the schematic stage of that genre and this can be attributed to the different rhetorical purposes of the sections.

In spite of the many studies conducted on the theme/rheme structure in different genres, very few studies, considering the number of articles published yearly, have been conducted on research articles in different registers (e.g., texts written in two settings) and disciplines. For example, Persian writers of English may transfer some elements unconsciously, and this may consequently affect the thematic structure of their texts in English. What is needed is a more thorough investigation of the theme and rheme and the role they play in textuality. Too little of the linguistic information that might be useful for non-native writers of English has yet filtered through to the academic system for non-native speakers. For that reason, the present research aims to locate the possible differences between English research articles of local and international journals with regard to theme selection and thematic progression patterns. It intends to highlight the thematic problems which local Persian writers of English might face in writing their articles, thus illustrating the need for informing local writers of English of the crucial role of theme/ rheme organization. As Borsinger (2003) argues, an awareness of this organization can become a powerful means for communicating results more effectively (p. 33). This research attempts to illustrate the possible similarities and differences between the writing of local and international writers regarding the theme choices and thematic

progression, hoping that changes in theme would indeed lead to articles being published internationally.

3. Methodology

3.1 Dataset: Description of Journals

The materials in this study consisted of a corpus of 32 articles: 16 from the *ELT Journal* and 16 from the *Roshd Journal*. The selection of these journals was motivated by the fact that *Roshd Foreign Language Journal* and *ELT Journal* are established scholarly journals which are widely accepted by experts in the field of ESL/EFL at local and international levels respectively. *Roshd* is a local, peer-reviewed journal published quarterly by the Ministry of Education in Iran. It targets English language teachers, students who major in teaching English, and all of those interested in the field of teaching foreign languages. It aims to help local teachers share and benefit from each other's experiences. This journal usually accepts articles in German, English, and Persian. It publishes predominantly MA thesis-driven articles jointly written by novice researchers and their supervisors.

Both journals provide a medium for discussion of the principles and practices of teaching and learning English language and are a forum for exchanging information among expert and novice members who wish to gain improved access to the wealth of ideas in the field of English language teaching.

3.2 Analytical Models

The models used for comparing and analyzing the dataset were Halliday (1985)'s categorization of theme and the revised model of Danes's (1974) thematic progression patterns proposed by McCabe (1999).

3.3 Halliday's Categorization of Theme

Halliday (1985, pp. 40-56) classifies themes into the following types:

1. Textual themes: conjunctions, conjunctive adjuncts, and continuatives,
2. Interpersonal themes: modal adjuncts, finite operators, and imperative *let's*,
3. Marked themes: adverbial groups, prepositional phrases, and complements which are nominal groups that are not functioning as subjects,
4. Unmarked themes: participants (subjects), extraposition (it), existential (there), thematic equatives (e.g., <u>what is important now</u> is to analyze the results), processes, and WH- elements in interrogatives,
5. Simple themes e.g., <u>I</u> went to London last week,
6. Multiple themes: textual^ topical marked, textual^ topical unmarked, interpersonal ^topical marked, interpersonal ^topical unmarked, textual^ interpersonal ^topical marked, textual^ interpersonal ^topical unmarked.

Following Martinez (2003), we opted for Halliday's categorization of theme on the ground that a) it relates theme to the ideational and interpersonal metafunctions and merges these two types of meaning in the clause initial position; b) it relates theme to the other subcategorizations of the systemic functional theory such as conjunctions and modal adjuncts; and c) through the concept of multiple themes, it provides both obligatory and optional elements (pp.108-109). Optional elements which include textual and interpersonal themes are crucial factors in understanding sentences from thematic perspectives; interpersonal meaning can be realized by means other than interpersonal theme (e.g., modals and hedged verbs or adverbs), and textual themes can express the logical relations explicitly and implicitly.

3.4 The Revised Model of Danes's TP Patterns

This study considered the revised model of Danes's (1974, as cited in Belmonte & McCabe, 1998) TP patterns proposed by McCabe (1999). Danes (1974, p. 114; cited in Belmonte & McCabe, 1998) identified four main patterns of thematic progression:

1) Constant thematic progression: The theme of one clause is derived from the theme of the previous clause;
2) Linear thematic progression: The theme of one clause is derived from the rheme of the previous clause;
3) Derived hyperthematic progression: Particular themes in subsequent clauses are derived from a hypertheme or from the same overriding theme;
4) Split rhematic progression: The rheme of the first clause is split into two items, each in turn being taken as a theme element in subsequent clauses.

According to McCabe (1999), there is a rather high percentage of clauses which does not fit into any of the TP patterns suggested by Danes (1974), since it seems that Danes employed a criterion for theme specification which accords more with the notion of "given" (p. 270). In this notion, theme necessarily links back either with a previous theme or rheme, while it is optimal for the writers to choose elements from the previous discourse to take as a point of departure for any given clause. Therefore, Danes' model needed to be modified in order to be used in other analyses which apply a different criterion for theme specification. McCabe (1999) added a fifth classification to Danes's, 5) split theme progression, where the theme of the first clause is split into two or more ideas, and these ideas are developed in the themes of subsequent clauses (p. 175).

McCabe (1999, p. 176) grouped these patterns into two overall types: a) theme progression including constant theme and split theme and b) rheme progression including simple linear and split rheme. She did not consider derived theme as a different sort of TP

since it may be related to preceding themes and rhemes through the same types of inference involved in simple linear or constant theme.

Following Martinez (2003, p. 108), the rationale for the selection of the above classifications was that they are more plausible, attestable, and complete than the other categorizations (e.g., Danes, 1974, as cited in Belmonte & McCabe, 1998) used for analyzing theme and TP patterns in texts. It is important to note that T-unit was adopted as the unit of analysis because this was recognized as the optimal unit for thematic progression in textual analysis.

3.5 Procedure

All the available volumes of the *ELT Journal* and the *Roshd Journal* published between the years 2002 and 2007 were collected. In order to narrow down the research scope, only experimental articles were considered. As a result, 32 articles (16 from each journal) were selected at random ensuring that the articles represented a variety of authors to control for personal writing style or preference. The *Roshd* authors were all Persian, representing researchers in the local EFL community whereas *ELT* authors portrayed a wider discourse community, with different linguistic backgrounds who found the chance to publish in ELT.

To ensure that the two groups are equal quantitatively, the articles excluding abstracts, footnotes, endnotes, reference lists, appendices, and texts associated with tables and graphics, either obtained directly from the electronic versions of the relevant journals or manually scanned and converted into Rich Text format, were saved on the computer and a word count was run on them. The corpus contained 100,960 words (50,644 for *Roshd Journal* and 50,316 for *ELT Journal*). The data were then carefully read and analyzed to calculate the frequency of theme types and thematic progression patterns in different rhetorical sections of the articles.

One problem with textual analysis is that there is always the danger of making mistakes in interpretation. To preclude the threat to reliability in the analysis, six articles from the corpus were also analyzed by an experienced researcher in applied linguistics and agreement was made on the method of analysis.

3.6 Rhetorical Specification

Initially the present study attempted to determine the different types of theme categorization and TP patterns in the four rhetorical sections of research articles (Introduction, Methodology, Results, and Discussion). But it was observed that a number of articles (in both journals) did not follow this standard format, and in fact their structures were different. Therefore, to avoid any possible wrong classification and to be consistent throughout this study, we decided to consider the results and discussion sections as one macro-part. In other words, the frequencies of theme types and TP patterns were counted and calculated in three rhetorical sections, namely introduction, methodology, and results and discussion.

3.7 Unit of Analysis

One of the primary considerations in the analysis of thematicity in texts is the unit of analysis as a basis for thematic identification. In the identification of T-unit, we encountered certain problems which are explained below.

According to McCabe (1999), "T-unit or independent conjoinable clause complex consists of an independent clause plus all of its hypotactically related clauses" (p. 73). The independent clauses can exist on their own as simple sentences or can string together using coordinating conjunctions. One of the major problems related to the coordinated independent clauses is ellipsis (North, 2005, p. 139). Following North (2005), in this study we considered the coordinated clauses with ellipted subjects as separate units provided that the clause still

contained a finite element, as in the following example, where the subject of the second clause is ellipted but the clause still contains the finite *were*.

(1) *ELT* 15: T-U 22: The students were allowed to work at their own pace,

T-U 23: *but were requested* to keep in regular contact.

If neither the subject nor the finite operator was present, the clause had no mood element, and it was regarded as an example of coordination at the level of group rather than clause. Note the following example.

(2) *Roshd* 9: T-U 212: Feedback will keep the dedicated teacher aware of the

learners' errors *and give* her ideas for tackling their learning problems.

Another problem occurred where an initial reporting verb might be followed by one or more clauses. McCabe (1999) states that the projected clauses following a reporting verb and introduced by *that* or having an ellipted subject are hypotactic clauses (p. 79). The clause that follows *that* in the example below is therefore regarded as hypotactic and cannot be assumed to be one T-unit but will constitute one single T-unit with the section that precedes it.

(3) *ELT* 3: T-U10: Schmitt states that the use of performed lexical chunks facilitates

greater fluency in speech production and aids the listeners.

4. Data Analysis

Considering the above points, we calculated the number of words and T-units in the three rhetorical sections of *ELT* and *Roshd* articles. The average number of words per clause in *ELT* and *Roshd* articles was 20.41 and 21.11 respectively. In both journals, the results and discussion sections included more T-units, which might be indicative of greater length of this section of articles. However, other differences could also be observed at this level. While 60% of T-units were found in the result and discussion sections of *ELT* articles, only 39% of T-units were included in this section of *Roshd* articles. A brief analysis of the introduction and

methodology sections also shows more differences across the two corpora. While these rhetorical parts were treated rather equally in *ELT* in terms of the number of T-units, Iranian-English article writers were more elaborate in the introduction, indicating that they treated this section almost close to the discussion and results sections in terms of the number of T-units. The results are shown in the following tables.

Table 1 Word Counts and T-Units in Different Rhetorical Sections of ELT Articles

ELT article	Word count				T-Units			Total
	I	M	R& D	Total	I	M	R&D	
ELT 1	441	776	2424	3641	17	29	106	152
ELT2	645	890	1422	2957	24	41	68	133
ELT3	1094	536	1272	2902	44	26	60	130
ELT4	184	726	2316	3226	7	32	122	161
ELT5	292	324	2574	3190	13	15	130	158
ELT6	1256	464	1609	3329	56	25	106	188
ELT7	179	708	2112	2999	6	41	110	157
ELT8	1471	536	1369	6676	69	20	66	155
ELT9	1313	806	695	2814	46	30	32	108
ELT10	949	779	1660	3388	40	48	85	173
ELT11	131	1043	1461	2635	5	52	76	133
ELT12	258	238	2807	3303	13	12	170	195
ELT13	1018	646	1346	3010	54	22	75	151
ELT14	583	503	1343	2429	26	26	53	105
ELT15	260	571	2689	3520	13	25	142	180
ELT16	808	444	2345	3597	36	24	127	187
Total	10882	9990	29444	50316	469	468	1528	2465

Columns: Numbers of words and T-units in the three rhetorical sections of ELT articles (introduction (I), methodology (M), and results and discussion (R&D))

Table 2 Word Counts and T-Units in Different Rhetorical Sections of Roshd Articles

Roshd Article	Word count				T-Units			Total
	I	M	R& D	Total	I	M	R&D	
Roshd1	861	531	753	2145	39	30	32	101
Roshd2	1827	879	657	3363	91	53	30	174
Roshd3	743	1118	1283	3144	36	52	55	143
Roshd4	1264	658	977	2899	46	33	42	121
Roshd5	908	547	2295	3751	39	35	132	206
Roshd6	755	1459	477	2691	31	80	22	133
Roshd7	1078	507	1675	3260	48	25	76	149
Roshd8	776	891	760	2427	38	44	31	113
Roshd9	1146	1047	1914	4107	56	67	87	210
Roshd10	1454	234	543	2231	80	12	23	115
Roshd11	820	313	1218	2351	30	16	51	97
Roshd12	819	600	956	2375	45	44	54	143
Roshd13	1118	644	3661	5423	47	31	145	223
Roshd14	2104	451	1327	3882	87	19	61	167
Roshd15	924	549	1775	3293	38	30	88	156
Roshd16	1838	836	628	3302	78	45	25	148
Total	18435	11311	20899	50644	829	616	954	2399

Columns: Numbers of words and T-units in the three rhetorical sections of Roshd articles (introduction (I), methodology (M), and results and discussion (R&D))

4.1 Theme Types in ELT Journal and Roshd Journal

The corpus was analyzed for the frequency of different types of theme. The total theme types were somewhat similar across the two journals. The number of textual themes was by far much greater than the number of interpersonal themes in both groups. This was compatible with Ghadessy (1999), McCabe (1999), and Whittaker's (1995) findings. Whittaker (1995) believes that this finding is not surprising since scientific writing tends to be impersonal and objective; moreover, the purpose of its writer is to persuade readers to read it (p. 109).

In the analysis of topical themes (Halliday, 1985, 1994) of both journals, it was found that the majority of topical themes were unmarked, implying that they occupied both the thematic and subject positions, once again similar to the conclusion drawn in Ghadessy (1999), McCabe (1999), and Whittaker's (1995) studies. The findings indicate the attention paid to grammatical subjects in academic articles, hence in the theme position.

Simple themes were substantially used in both journals. The frequency of multiple themes was 569 (24%) and 549 (23%) in *ELT* and *Roshd* respectively. The analysis showed that multiple themes were outstandingly unmarked in both journals (72% in ELT and 82% in Roshd), though not with the consistency observed in Gomez's (1994) study of news reports in which she found that 99% of multiple themes were unmarked. The difference is likely to be genre-related. The different types of theme (except interpersonal theme in *Roshd Journal*) concentrated in the results and discussion sections of the articles. This may imply the rhetorical goals of the results and discussion section where writers make their claims, suggestions, and recommendations, and explore implications tied directly to their study.

In the following sections, the different types of theme are explained separately. The data presented in Table 3 shows the total frequency and percentage of theme types in *ELT* and *Roshd* articles in general without discriminating their rhetorical sections.

Table 3 Frequency and Percentage of Themes in ELT and Roshd Journals

	Textual (%)	Interpersonal (%)	Marked (%)	Unmarked (%)	Simple (%)	Multiple (%)
ELT	565(23)	95(4)	677(27)	1788(73)	1869(76)	596(24)
Roshd	538(22)	60(2.5)	579(24)	1820(76)	1850(77)	549(23)

4.2 Textual Themes

Textual themes were identical in both corpora quantitatively, in line with McCabe's (1999) results. In the analysis of history texts, she found 23.40% textual themes in English and 23.91% in Spanish texts, much more than the findings by Whittaker (1995), who reported only 15% textual themes in the analysis of linguistics and economics articles.

In *ELT* articles, coordinated conjunctions were used more frequently than conjunctive adjuncts. The situation was almost opposite in *Roshd* articles. In other words, while international writers tended towards conjunctions as textual themes, local writers leaned towards conjunctive adjuncts. This is very similar to McCabe's (1999) finding that the most

frequent types of textual theme used in English and Spanish history textbooks were conjunctions in the former and conjunctive adjuncts in the latter (p. 221).

It was also revealed that the conjunction type which had the highest frequency in the three rhetorical sections of both groups was *and*. Moreover, the most frequent conjunctive adjuncts in the rhetorical sections of *ELT* articles were adversatives in the introduction and results and discussion sections. In the method section, since the frequency of conjunctive adjuncts was low, and at the same time, their variety was high, we could not decide on the one with the highest frequency. Some familiar examples included additives (e.g., *in addition, moreover, furthermore*) and summatives (e.g., *in sum, all in all, in a nut shell*). In the case of *Roshd* articles, the more frequent conjunctive adjuncts were temporal conjunctives (e.g., *first*, *finally*…) in the introduction and method sections and adversatives in the results and discussion sections.

The continuative was not utilized except by ELT researchers in the results and discussion sections, although its frequency was low (6 cases or 0.24%). This complied with McCabe's (1999) study in which she found no continuatives in the English corpus and only two instances in the Spanish corpus. The findings are illustrated in Tables 4 and 5 below.

Table 4 Frequency of Textual Themes in Three Rhetorical Sections of ELT Journal

Textual types	I (%)	M (%)	R& D (%)	Total
conjunction	45(1.82)	61(2.47)	207(8.4)	313(12.69)
conjunctive	56(2.27)	26(1.05)	164(6.65)	246(9.97)
continuative	0	0	6 (0.24)	6(0.24)
Total	101(4.09)	87(3.52)	377(15.29)	565(22.9)

Columns: Rhetorical sections of ELT journals (introduction, methodology, and results and discussion)

Table 5 Frequency of Textual Themes in Three Rhetorical Sections of Roshd Journal

Textual types	I (%)	M (%)	R& D (%)	Total (%)
conjunction	67(2.8)	69(2.87)	84(3.5)	220(9.17)
conjunctive	104(4.33)	57(2.37)	157(6.55)	318(13.25)
continuative	0	0	0	0
Total	171(7.12)	126(5.25)	241(10.05)	538(22.42)

Columns: Rhetorical sections of Roshd Journals (introduction, methodology, and results and discussion

4.3 Interpersonal Themes

The two groups were analyzed in terms of the frequency of interpersonal themes, and a relative similarity was found across the two journals. This finding was compatible with Martinez's (2003) study of biology research articles in which the interpersonal theme occurred less frequently, attributing this low occurrence to the formality of the register.

Results, however, were in contrast to Whittaker's (1995) analysis of academic articles and Ghadessy's (1999) sport commentaries analysis. They found more interpersonal themes in the texts. Martinez (2003) states that the difference with Whittaker's finding suggests a high degree of impersonality in the academic articles and a greater degree of commitment to the truthfulness of what they are claiming, whereas the difference with Ghadessy's study can be explained in terms of generic differences; that is, sport commentaries are more conversational (p.111). Results are illustrated in Tables 6 and 7 below.

Table 6 Interpersonal Themes in Rhetorical Sections of ELT Articles

Interpersonal types	I (%)	M (%)	R& D (%)	Total
Modal	5(0.2)	5(0.2)	43(1.75)	53(2.15)
Finite	13 (0.53)	10(0.4)	17(0.69)	40(1.62)
Let's	1(0.04)	0	1(0.04)	2(0.08)
Total	19(0.77)	15(0.6)	61(2.47)	95(3.85)

Table 7 Interpersonal Themes in Rhetorical Sections of Roshd Articles

Interpersonal types	I (%)	M (%)	R& D (%)	Total
Modal	11(0.46)	1(0.04)	17(0.71)	29(1.21)
Finite	19 (0.8)	6(0.25)	6(0.25)	31(1.29)
Let's	0	0	0	0
Total	30(1.25)	7(0.29)	23(0.96)	60(2.5)

As the above tables show, the interpersonal themes were more frequent in *ELT* than *Roshd* articles. The most frequent types of interpersonal themes were modal adjuncts in *ELT* and finite operators in *Roshd* articles. The modal adjuncts were mainly concentrated in the results and discussion sections of the two journals. The most frequent type of modal adjunct in the results and discussion sections was validities (e.g., *generally, frankly, perhaps, hopefully*). In the other sections of the articles, although the number of different types of interpersonal theme was very low, their varieties were high, and so a decision upon the most frequent interpersonal themes did not offer a clear picture of theme distribution and scarcity.

4.4 Marked Themes

Marked themes in general were used in a similar way in both groups. In both journals, the proportion of marked themes was higher in the results and discussion sections, but this tendency was greater in *ELT* articles. The findings are in agreement with Whittaker's (1995)

and Martinez's (2003) findings. They attribute this greater proportion of the marked theme to the argumentative nature of the paragraphs and sections in which this kind of theme is used.

One marked difference was while the introduction and results and discussion sections were diverse in terms of the marked theme in *ELT* articles, the difference was very small in *Roshd* articles. The number of marked themes in the results and discussion sections of *ELT* was three times of the introduction section, indicating the elaborate nature of the results and discussion sections wherein international writers argue the points they make in more detail. Table 8 shows the frequency and percentage of marked themes in the different rhetorical sections of *ELT* and *Roshd* articles.

Table 8 Marked Themes in Three Rhetorical Sections of ELT and Roshd Journals

Journals	I (%)	M (%)	R&D (%)	Total (%)
ELT journal	131(5.31)	144(5.85)	402(16.30)	677(27.46)
Roshd journal	207(8.63)	129(5.38)	243(10.12)	579(24.13)

Columns: Rhetorical sections of ELT and Roshd journals (introduction, methodology, and results and discussion)

4.5 Unmarked Themes

The total percentage of unmarked themes was similar in both journals. The most frequent type of unmarked theme was *participant* in both corpora. *Participant* was mainly concentrated in the results and discussion sections of the two groups, close to Ghadessy (1999) and Whittaker's (1995) findings. In the analysis of their data, they found the greatest majority of topical themes (Halliday, 1985, 1994) were the subjects or participants. Other types of unmarked theme were either rarely used or they were not utilized in the data at all.

One significant difference was the way in which unmarked themes were distributed in the introduction and results and discussion sections of both journals. In *ELT* articles, the difference between the introduction and results and discussion was very significant. That is, while 339 cases or 13.75% of unmarked themes were employed in the introduction section,

the number of unmarked themes in the results and discussion section was 1125 or 45.64%. In *Roshd* articles, on the other hand, the difference was very minimal. The frequency of different types of unmarked theme in *ELT* and *Roshd* journals is depicted in Tables 9 and 10 below.

Table 9 Unmarked Themes in Different Rhetorical Sections of ELT Journals

Unmarked themes	I (%)	M (%)	R&D (%)	Total
Participants	310(12.57)	301(12.21)	1056(42.84)	1667(67.62)
Extraposition	12(0.48)	9(0.36)	36(1.46)	57(2.3)
Existential	6(0.24)	2(0.08)	13(0.52)	21(0.85)
Equative	1(0.04)	0	10(0.4)	11(0.44)
Process	1 (0.04)	0	4(0.16)	5(0.2)
Wh-	9 (0.36)	12(0.49)	6 (0.24)	27(1.09)
Total	339(13.75)	324(13.14)	1125(45.64)	1788(72.53)

Table 10 Unmarked Themes in Different Rhetorical Sections of Roshd Journals

Unmarked themes	I (%)	M (%)	R&D (%)	Total
Participants	581(24.22)	471(19.63)	645(26.88)	1697(70.73)
Extraposition	9(0.37)	5(0.21)	45(1.88)	59(2.46)
Existential	24(1)	8(0.33)	12(0.5)	44(1.83)
Equative	1(0.04)	0	5(0.20)	6(0.25)
Process	0	0	1(0.04)	1(0.04)
Wh-	6 (0.25)	3(0.12)	4 (0.16)	13(0.54)
Total	622 (25.92)	487(20.3)	711 (29.63)	1820(75.86)

4.6 Simple Themes

Comparison of simple themes in different rhetorical sections of the articles revealed that the unmarked simple themes were dominant in the *ELT* and *Roshd* articles (75% and 83% of all themes respectively). They were mainly found in the results and discussion sections. This is

somewhat in the line with Martinez's (2003) findings. In the analysis of the method and discussion sections of biology articles, she reported a high percentage of unmarked simple themes in both sections, and she argues that unmarked simple themes have an important function, giving continuity to the texts (p. 118).

One area of variation was the distribution of simple themes across the different sections in both groups. In *ELT* articles, about an equal number of simple themes appeared in the introduction and method sections, while this theme was three times greater in the results and discussion sections than each of the other sections. In *Roshd* articles, simple themes were almost equally distributed across the introduction and results and discussion sections while the method section only had a smaller number of simple themes.. The results are shown in Table 11 below.

Table 11 Simple Themes in Different Rhetorical Sections of ELT Journal

Journal	I (%)	M (%)	R&D(%)	Total
ELT	364(14.77)	381(15.45)	1124(45.60)	1869(75.82)
Roshd	647(26.96)	491(20.46)	712(29.69)	1850(77.11)

Columns: Rhetorical sections of ELT and Roshd journals (introduction, methodology, and results and discussion)

4.7 Multiple Themes

Like the other types of themes, the total frequency of multiple themes was similar between the two journals. The most frequent type of multiple themes was *textual ^ unmarked* themes. Though both groups tended to use *textual^ unmarked* multiple theme in the results and discussion sections, the tendency was greater in *ELT* articles. Next in rank was *textual ^ marked* multiple themes in both journals. The frequency of multiple themes containing interpersonal themes was low. This seems to confirm Martinez's (2003) study in which she found that the percentage of multiple themes was four times as high as that in the method

section. She believed that this thematic choice illustrates the author's rhetorical effort to persuade the reader. The findings are summarized in Tables 12 and 13.

Table 12 Multiple Themes in Different Rhetorical Sections of ELT Journal

Multiple themes	I (%)	M (%)	R&D (%)	Total
Text^ unmarked	58(2.36)	52(2.1)	253(10.26)	363(14.72)
Text^ marked	31(1.25)	21(0.85)	87(3.52)	139 (5.63)
Inter^ unmarked	8(0.32)	2(0.08)	45(1.82)	55(2.23)
Inter^ unmarked	0	1(0.04)	3(0.12)	4(0.16)
Text^inter^unmarked	5(0.2)	1(0.04)	6(0.24)	12(0.48)
Text^inter^marked	4(0.16)	10(0.4)	9(0.36)	23(0.93)
Total	106(4.3)	87(3.52)	403(16.34)	596(24.17)

Columns: Rhetorical sections of ELT and Roshd journals (introduction, methodology, and results and discussion)

Table 13 Multiple Theme in Different Rhetorical Sections of Roshd Journal

Multiple themes	I (%)	M (%)	R&D(%)	Total
Text^ unmarked	123(5.13)	97(4.04)	199(8.29)	419(17.46)
Text^ marked	30(1.25)	21(0.88)	27(1.12)	78 (3.25)
Inter^ unmarked	17(0.71)	2(0.08)	6(0.25)	25(1.04)
Inter^ unmarked	4(0.16)	0	0	4(0.16)
Text^inter^unmarked	2(0.08)	0	6(0.25)	8(0.33)
Text^inter^marked	6(0.25)	5(0.2)	4(0.17)	15(0.62)
Total	182(7.59)	125(5.21)	242(10.09)	549(22.89)

Columns: Rhetorical sections of ELT and Roshd journals (introduction, methodology, and results and discussion)

Frequencies per se might not tell us much about the scale of difference between the theme types under the study. Therefore, a *chi-square* analysis was utilized in order to see if the existing difference was statistically meaningful. The results are shown in Table 14. Since the values of c*hi*- square obtained for the different types of themes were less than 3.84, the

differences were not statistically meaningful. In other words, local and international articles did not exhibit a marked difference in utilizing themes.

Table 14 Results of Chi-square Test

Theme types	Total words	Total themes	Percent	x^2	P
Textual in ELT	2465	565	22.90		
Textual in Roshd	2399	536	22.42	0.661	0.080
Interpersonal in ELT	2465	95	3.85		
Interpersonal in Roshd	2399	60	2.50	3.063	0.080
Marked in ELT	2465	677	27.46		
Marked in Roshd	2399	579	24.13	0.076	0.677
Ummarked in ELT	2465	1788	72.53		
Unmarked in Roshd	2399	1820	75.86	0.060	0.806
Simple in ELT	2465	1860	75.82		
Simple in Roshd	399	1850	77.11	0.007	0.084
Multiple in ELT	2465	596	24.17		
Multiple in Roshd	2399	549	22.88	0.021	0.884

4.8 Thematic Progression Patterns

All T-units in the corpus were analyzed to determine the different patterns of thematic progression: namely, constant, linear, split theme, and split rheme. According to McCabe (1999), the relationships are determined on the basis of the presence of different cohesive

devices, such as repetition, synonymy, antonymy, hyponymy, pronouns, substitutions, ellipsis, and paraphrasing (p. 176).

In this study, we followed North's (2005) procedure for identifying links by looking for the first most constituent with a topical link with the proceeding main clause. If the left most item identified as a link was a modifier, then the whole constituent was examined for a link to the preceding text.

Verb-initial constructions with ellipsis were analyzed as forming a part in a simple linear progression chain. According to North (2005), these constructions have come from the previous rhemes, so they have a linear progression provided that the subject for the verb is the same and the verbs have the same inflection for the tense. Note the following two T-units where the finite and the verb inflection are the same.

(4) ELT T-U 37: The groups were coded as control and experimental groups

T-U 38: and *were supposed* to receive routine method and treatment respectively.

Following McCabe (1999), the themes for which the reader has to go back more than two clauses to find a previously mentioned concept were not considered as part of thematic progression patterns. In other words, if links were established over three or more clauses, these links were not considered as part of thematic progression patterns.

The last point to be mentioned here is that some of the themes in the corpus could not be analyzed in terms of thematic progression patterns; in other words, they did not fit into the above configurations. Following McCabe (1999), we labeled them as *key themes, summary themes, pervious theme, back themes, related themes,* and *new themes.*

The other types of themes which lay outside the thematic bonds were *metatextual* themes (referring to text itself e.g., *this study*), *extralinguistic* themes (e.g., *you, we*), *existential* themes (*there*), *extraposition* (*it*), *adverbial* themes (*temporal*), *process, wh-*

elements in interrogative clauses, thematic equatives, general words (e.g., *the problem* or *one*), and *examples*. The results are shown in Table 15.

Table 15 Thematic Progression Patterns in ELT and Roshd Journals

Journals	Linear (%)	Constant (%)	Split theme (%)	Split rheme (%)	Total (%)	Peripheral (%)
ELT	925(38)	662(27)	11(0.44)	15(0.6)	1613(66)	852(34.56)
Roshd	824(34.34)	803(33.47)	3(0.12)	37(1.54)	1667(69.48)	732(30.51)

As shown in Table 15, the tendency for split themes and split rhemes was very low in both journals; both groups also preferred linear themes over constant themes. However, the linear theme was more frequent in *ELT* articles. In *ELT* articles, the results are in agreement with McCabe's (1999) study in which linear chains occurred more than constant chains in history texts, attributing this to the analytical or explicit nature of history texts. This was expected due to the pedagogical purpose of the texts and asymmetric relationships between the writer and the reader (p. 23).

Peripheral themes or themes outside the thematic progression patterns were considerably used in both corpora. The most common type was *back* theme in both journals (9%) followed by *previous* and *metatextual* themes (McCabe, 1999).

In the next step, the frequencies and percentages of different patterns of thematic progression were calculated in the three rhetorical sections of the corpus. The results are illustrated in Tables 16 and 17 below.

Table 16 Thematic Progression Patterns in ELT Rhetorical Sections

R S	Linear (%)	Constant (%)	Split theme (%)	Split rheme (%)	Total (%)	Peripheral (%)	Total (%)
I	177(7.18)	92(3.73)	o	4(0.16)	273(11.07)	196(7.95)	469(19.02)
M	182(7.38)	133(5.39)	6(0.24)	0	321(13.02)	147(5.96)	468(18.98)
R&D	566(23)	437(17.72)	5(0.2)	11(0.44)	1019(41.36)	509(20.64)	1528(62)

Table 17 Thematic Progression Patterns in Roshd Rhetorical Sections

RS	Linear (%)	Constant (%)	Split theme (%)	Split rheme (%)	Total (%)	Peripheral (%)	Total (%)
I	281(11.71)	303(12.63)	o	11(0.46)	595(24.8)	234(9.75)	829(34.55)
M	232(9.67)	192(8)	3(0.12)	19(0.79)	446(18.59)	170(7.08)	616(25.67)
R&D	311(12.96)	308(12.83)	0	7(0.3)	626(26.09)	328(13.67)	954(39.76)

In the introduction section, the total percentage of thematic strings in *ELT* articles was less than half of the patterns in *Roshd* articles. Constant theme (Belmonte & McCabe, 1998) was more frequent in *Roshd* articles because, as Belmonte and McCabe (1998) argue, the writer tends not to go in depth on the idea introduced in the rheme of the text. Linear and peripheral themes were also more frequent in *Roshd* Journal. This can be attributed to the greater number of T-units in the introduction sections of *Roshd* articles. The most frequent type of peripheral theme was *back* in the introduction section of both journals. The other more frequent types were *metatextua*l in *Roshd* and *related* (McCabe, 1999) in *ELT* articles.

In the methodology section, like the introduction, linear and constant themes were more frequent in *Roshd* articles. The most frequent type of peripheral theme in this section was *back* followed by *metatextual* (McCabe, 1999).

In the results and discussion sections, there were large differences in the frequency of linear and constant chains in both journals. These chains were more frequent in *ELT* articles which indicates that the results and discussion sections were more elaborate and argumentative than *Roshd* articles. Like the other sections, *back* theme was the most frequent type of peripheral themes. The other more frequent types were *metatextual* in the case of *Roshd* and *previous* theme in *ELT* articles. In order to compare the frequencies in both groups, the *chi-square* analysis was used. The result is shown in Table 18.

Table 18 Results of Chi-square Test in ELT and Roshd Journals

Corpus	Total words	Total TP	Percent	x^2	P
ELT	2465	1613	65	0. 730	0.119
Roshd	2399	1667	69		

-Critical value was 3.84.
-Level of significance was set at 0.5.

In addition to this comparison, *chi-square* was carried out to compare the frequency of thematic bonds in different rhetorical sections of the articles. The results are illustrated in Table 19.

Table 19 Results of Chi-Square across Different Rhetorical Sections

Corpus	Total words	Total words in RS	Total thematic progression	Percent	x^2	P
ELT I	2465	469	273	11.07	5.444	0.020
Roshd I	2399	829	595	24.8		
ELT M	2465	468	321	13.02	1.125	0.289
Roshd M	2399	616	446	18.59		
ELT R&D	2465	1528	1019	41.33	3.958	0.049
Roshd R&D	2399	954	626	26.09		

While the difference between the overall frequencies of thematic chains in both corpora was not statistically significant (Table 18), the results indicated significant differences in the introduction (x^2=5.444) and results and discussion (x^2=3.958) sections of the *ELT* and *Roshd* articles since the values of *chi*-square obtained were greater than the critical value (3.84).

5. Discussion

Analysis of the data reveals interesting points vis-a-vis the research focus addressing the different types of theme and thematic progression patterns, which are discussed below in turn.

5.1 Theme Types

The similarities in using theme types in both journals can be explained in terms of genre. It is believed that the texts belonging to the same genre present a similar contextual configuration (Halliday, 1985 & 1994; Swales, 1990); that is, they illustrate common characteristics in linguistic elements related to the field, mode, and tenor of discourse. According to Halliday and Hasan (1985), the field of discourse recognized in the culture can predict the experiential meaning. Tenor of discourse or the actors, involved in the creation of text, can predict interpersonal meaning; and mode of discourse, or particular functions assigned to language and rhetorical channels, can predict textual meanings (p. 46).

Similarities in the discourse characteristics engender similarities in textual choices, which are reflected in the thematic choices preferred, since the field, tenor, and mode can be realized in topical, interpersonal, and textual themes respectively. Theme is one of the textual choices that provides interesting insights into establishing similarities within and between genres. It is worth noting that this study, in the line with Ghadessy (1999), McCabe (1999), and Whittaker (1995), proves that an analysis of different types of theme reveals crucial characteristics of academic articles.

The findings obtained from the analysis of different rhetorical sections of *ELT* and *Roshd* articles confirm that the frequencies of textual, marked, and multiple themes were higher in the results and discussion sections than the method section of both journals. According to Martinez (2003), the discussion section is the most abstract and argumentative part of any article, whereas the method section is characterized by a factual and expository nature (p.105). So based on the findings of this study, textual and marked themes are associated with the argumentative nature of results and discussion sections.

In addition to the impersonal and factual tone of research articles, the low percentage of interpersonal themes might reflect the lack of symmetrical relationships between the reader

and the writer. The writer, as McCabe (1999) states, has information that the reader does not, and thus he is not attempting to persuade the reader to agree with his statements, since he presents the bulk of his statements as factual (p. 228). Overall, the greater number of interpersonal themes in *ELT* reveals the greater tendency of international articles towards reader-directness and reader- friendliness of their writing, which eventually softens the textual effect.

The greater inclusion of finite operators in the results and discussion sections of *ELT* articles could imply that the writers of *ELT* articles were aware of the function of finite operators since a finite operator, as Halliday (1994) states, makes the proposition finite, down to earth, and arguable, and this is achieved by relating the proposition to the context in which it is used (p. 75). So finite operators can help the argumentative nature of the results and discussion sections. Finite operators were, however, kept in low profile in the results and discussion sections of *Roshd* articles, attesting to the fact that, where more argumentation is required, local writers fail to provide sufficient explanation of their findings.

Moreover, the significant differences in the way that marked and unmarked themes were utilized in the introduction and results and discussion sections of both journals can be attributed to the distribution of T-units in these rhetorical sections. The longer results and discussion sections in *ELT* articles result in more frequent marked and unmarked themes, indicating that international writers argue the points they make in more detail. While the longer introduction in *Roshd* articles proves the descriptive and narrative nature of these articles, indicating that local writers paid more attention to description rather than argumentation.

The argumentative nature of results and discussion sections in *ELT* articles is also portrayed in utilizing *Textual ^ marked* and *Textual ^ unmarked multiple* themes. In other words, there is a close relationship between the distributions of different types of themes and

the way texts are perceived as argumentative. The results and discussion sections of *ELT* articles are thematically realized as *Textual* ^ *marked* and *unmarked*.

5.2 Thematic Progression Patterns

In compliance with the results of theme types, the diversity in using constant and linear TPs in the results and discussion sections might reveal the argumentative nature of ELT articles while about-equal treatment of these TPs in the introduction and results and discussion sections of *Roshd* articles might signal the inability of non-native writers to recognize the rhetorical function of these sub-genres. There is agreement among researchers, such as Belmonte and McCabe (1998), McCabe (1999), and Wang (2007), that academic texts present complex arguments in which an idea in the clause is expressed and explained in the successive clauses. These arguments help the reader to be aware of where the information has come from and where it is going, thus creating cohesion in the texts. Because of this argumentative nature, an academic text needs to have a high frequency of linear links which, as McCabe (1999) states, establish relationships implying cause and effect. Linear links also develop ideas and improve cohesion between sentences in a paragraph (p. 211).

Argumentative texts are characterized by high proportions of linear and constant thematic progression chains. The discussion section is the most abstract and argumentative section of academic articles where the results obtained must be evaluated and interpreted in relation to the problems, and the writer should try to persuade the reader to accept the relevance of the findings to other findings. Results attest to the local writers' unawareness of the rhetorical function of this section as the difference between the introduction and the results and discussion sections was thematically minimal. However, this was an area of stark contrast between local and international articles. According to McCabe (1999), the high proportion of constant theme in an academic article creates a static text in which an event,

item, or person is presented to the reader without a fuller picture of that event, item, or person or without providing any connection between the phenomena (pp. 202- 203). In other words, the over-use of constant chain yields simplistic, repetitive, and redundant paragraphs which may result from the continuous use of the same topic. It presents a list of events without expanding the information introduced in the theme, and the topic does not develop fully in the rheme. Nevertheless, the results of the present study do not confirm McCabe's thematic distinction made between article sections, either due to the disciplinary differences or due to our decision in merging the final two sections of articles. Future studies might take into account articles with separate results and discussion sections so that more logical judgment can be made on the thematic nature of these article sections.

The longer introduction parts in *Roshd* articles were due to longer literature reviews associated with them. This was depicted in the number of T-units in the introduction sections of both journals. In *Roshd* articles, the number of T-units almost doubled the number in *ELT* (829 or 35% in *Roshd* and 469 or 19% in *ELT*). One of the main features of the literature review is that it is associated with narration and description. Thus local writers overrelied on constant chains in order to establish backgrounds for their own study.

Concerning the generic nature of journal articles, *ELT* and *Roshd* articles share the features (common goals, mechanisms of interaction, specific vocabulary, and community members) mentioned by Swales (1990). In addition to the same form and rhetorical sections, there are similarities in terms of their audience, content, and purpose which result in similar frequencies of theme types and even the general similarity in the case of thematic progression patterns.

As Ghadessy (1995) states, the understanding of how texts are created would be much poorer without the concept of theme and thematic progression patterns (p. 159). The writers must be trained to be sensitive to what they write, and how they give their information in the

clauses and also in the texts. It is necessary to develop an awareness about different thematic options used by writers who succeed in the publication of their articles in international journals. Analyzing the theme structures of these articles can enhance the writing ability of local writers and establish semantic coherence in their articles. Finally, this study can shed light into the importance of theme in genre analysis. According to Lores (2004), thematic analysis can be applied as a useful tool for exploring genres since it is significant as part of a writer's available linguistic resources and can guide the writer to produce effective local and global discourse (p. 98). The findings of this research may help the writers to construct texts which are appropriate for the genre in question, and an awareness of the way theme choices are realized in the genre of academic articles can uncover some sources of difficulties that the writers encounter in writing their articles.

The study envisaged the usefulness of the concept of theme in describing the initial part of the clause in an academic context. Clearly our decision concerning the nature of the thematic structure of applied linguistic articles is by no means complete or comprehensive enough to account for all potential settings. Further studies are then needed to make convincing decisions in this regard. With the large number of articles published in different fields, definitely, more studies are needed to offer a pronounced picture of the thematic structure of local and international journal articles.

References

Baltazani, M. (2006). Intonation and pragmatic in interpretation of negation in Greek. *Journal of Pragmatics, 38*, 1658-1676.

Belmonte, I. A., & McCabe, A. (1998). Theme- Rheme patterns in L2 writing. *Didactica, 10*, 13- 31.

Borsinger, A. M. (2003). A comparison of the thematic options in novice and experts research writings. *Studios de Lingustica Aplicada Julio*, *37*(21). 37-51.

Daneš, F. (1974). Functional sentence perspective and the organization of the text. In F. Daneš (Ed.), *Papers on Functional Sentence Perspective* (pp. 106-128). Prague: Academia.

Fries, P. H. (1983). On the status of theme in English: Arguments from discourse. *Forum Linguisticum, 6*(1), 1-38.

Fries, P. H. (1995). A personal view of theme. In M. Ghadessy (1995). In M. Ghadessy (Ed.), *Thematic development in English texts* (pp. 1-19). London: Printer.

Ghadessy, M. (1995). Thematic development and its relationship to registers and genres. In M. Ghadessy (Ed.), *Thematic development in English texts* (pp. 129-146). London: Printer.

Ghadessy, M. (1999). Thematic organization in academic article abstract. *Estudious Lingleses de la Universi Dad Complutense, 7*, 141-161.

Gomez, M. (1994). Theme and textual organization. *Journal of Words, 45*(3), 293-305.

Halliday, M. A. K. (1985). *An introduction to functional grammar*. London: Edward Arnold.

Halliday, M. A. K. (1994). *An introduction to functional grammar*. London: Edward Arnold.

Halliday, M. A. K., & Hasan, R. (1985). *Language, context, and text: Aspects of language in social- semiotic perspective*. Oxford: OUP.

Jalilifar, A. R. (2009) Thematic development in English and translated academic texts. *Journal of Language and Translation, 10*(1), 81-111.

Ji, S. (2002). Identifying episode transitions. *Journal of Pragmatics*. *34*, 1257-1271.

Lores, R. (2004). RA abstracts: From rhetorical structure to thematic organization. *English for Specific Purposes, 23,* 280-302.

Martinez, I. (2003). Aspects of theme in method and discussion sections of biology journals articles in English. *Journal of English for Academic Purposes, 2*, 103-123.

Matthiessen, C., & Halliday, M. A. K. (1997). Systemic functional grammar: A first step into the theory. Retrieved September 5, 2006, from http://www.SFGintro(peng chapter) New.com

McCabe, A. (1999). Theme and thematic patterns in Spanish and English history texts. Unpublished Ph.D. thesis. Retrieved April 22, 2006, from: http:www.wag soft.com/systemics/archieve/ McCabe.phd.

McCabe, A. & Belmonte, I. A. (2001). Theme, transitivity, and cognitive representation in Spanish and English written texts, *Clac, 7*, 1-15.

McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge: CUP.

North, S. (2005). Emergent disciplinarity in an interdisciplinary course: Theme use in under-graduate essays in the history of science. Unpublished PhD Thesis. The Open University

Paltridge, B. (2006). *Discourse analysis.* London: Continuum.

Swales, J. M. (1990). *Genre analysis: English in academic and research setting.* Cambridge: CUP.

Wang, I. (2007). Theme and rheme in the thematic organization of texts: Implications for teaching academic writing. *Asian EFL Journal*, *9*(1), 164-176.

Whittaker, R. (1995). Theme, processes, and the realization of meaning in academic articles. In M. Ghadessy (1995). *Thematic development in English texts* (pp. 105-128). London: Printer.

The Asian ESP Journal

Citation Practices in TESL Research Articles: A Comparative Study

Chun-Chun Yeh

Biodata

Chun-Chun Yeh is Associate Professor in the Department of Foreign Languages and Literature, National Chung Cheng University, Taiwan. Her research interests include English for academic purposes, EFL writing instruction, and translation teaching.

Abstract

In the past two decades, there has been a steadily increasing number of studies investigating different aspects of research writing. Among the discourse elements in research writing, "citation," reference to previous literature, is argued to be a central feature that serves to persuade the reader of the soundness and novelty of one's own work. The current corpus-based study uses two corpora of 40 research articles selected from two TESL journals in order to explore and compare the similarities and differences between international and Taiwanese scholars' citation practices. Results show that the two corpora exhibit a number of similarities, such as a preference for summary/paraphrase and generalization over direct quotations; and an abundant use of citations in introductions. However, the analysis also found that Taiwanese scholars use fewer citations than their international counterparts. They use more dated and fewer recent references. In addition, they tend to avoid direct quotations.

Alongside these general tendencies, individual differences are also found among research texts in the same corpus.

Keywords: applied linguistics; citation; English for academic purposes (EAP); research articles

1. Research on citation use in research articles

In the past decades, there has been a steadily increasing number of studies investigating different aspects of research writing: for example, use of personal pronouns (e.g., Kuo, 1999), evaluative language use (e.g., Charles, 2003), hedging (e.g., Salager-Meyer, 1994), and various other features. These studies have shown that research writing is not merely plain reporting of scientific facts, but also a sociorhetorical act that involves careful consideration of social interaction within the discourse community (Bazerman, 1988; Myers, 1990). In order to gain acceptance in the discourse community, authors of research texts have to claim relevance for their own work by commenting on previous studies and identifying a gap in the research tradition (Swales, 1990). Citation, or reference to another's work, is therefore an important academic feature among the many devices that help achieve an author's communicative purposes.

Theories of citation have been proposed and debated (Cronin, 1998; Leydesdorff, 1998). Among them, two, both from a rhetorical point of view, are perhaps the most relevant here. In the first theory, citation has been argued to be a tool of persuasion, used to give greater authority to one's own text (Gilbert, 1977). In the other, Swales (1990) proposes that the purpose of citation is to create "a personal research space," by pointing out what has and has not been done, in order to prepare a space for new research. Both theories show that appropriate and strategic use of citation is an essential skill for academic researchers.

As a central feature in research writing, citation has received increasing attention in academia. Apart from contributions from disciplines like sociology of science and information science (see White, 2004), applied linguists have also examined various aspects of citation, such as integral vs. non-integral structures (Charles, 2006; Hyland, 1999; Swales, 1990), use of reporting verbs (Hawes & Thomas, 1997; Shaw, 1992; Thompson & Ye, 1991), self-citation (Harwood, 2005; Hyland, 2001), citation use across disciplines (Hyland, 1999), and cross-cultural use of citation (Bloch & Chi, 1995; Taylor & Chen, 1991). These studies illustrate how authors in various discourse communities conceptualize research issues and co-construct knowledge by citing previous work strategically. Although these studies have contributed to our understanding of this important feature of research writing, we know little about how experienced researchers writing in a second language (English) cite in English academic writing. To shed light on the issue, this paper will explore and compare citation use in 40 research articles (RAs) written by international and Taiwanese scholars. More specifically, it will examine various aspects of citation use—overall citation frequency, frequency in RA rhetorical sections, citation surface forms, presentation of cited information, and citation dates—by two groups of researchers working in the same field: teaching English as a second language (TESL), one group referred to as international scholars, the other, Taiwanese applied linguists.

2. Citation features in RAs

Recent years have seen a continuous progress in citation use in RAs. For the purpose of this paper, this section will review studies on citation frequency, spread of citations in RA rhetorical sections, citation forms, cross-disciplinary and cross-cultural differences in citation use, with a special emphasis on Chinese scholars' citation behavior.

2.1. Citation frequency in RAs

Research on the history of the RA has demonstrated that there has been a dramatic rise in the number of references over the last hundred years. Bazerman's (1988) study of experimental reports in *Physical Review* between 1893 and 1980 shows that researchers increasingly sought to ground their studies in the literature of the field. His study found that, after the 1960s, the number of references per article increased steadily, which in turn contributed to a rise in article length. In the early years, references generally did not refer to specific findings or "identify a specific relation to the current work" (p. 164). Furthermore, they sometimes were undated and occurred only in the beginning section. The later decades of the past century, however, saw a change in RA citation patterns. Researchers began to use more references and to integrate references more closely into the argument. References are now spread throughout the RA rather than appearing only in the beginning part. Similarly, Salager-Meyer's (1999) diachronic study of 162 medical RAs published between 1810 and 1995 shows a sharp increase in the number of references: 1.6 per article in the 1810–1849 block vs. 16.4 in the 1950–1995 block. Analyzing 80 contemporary RAs from eight disciplines, Hyland (1999) reports an average of 104 citations in sociology RAs and 75.3 in applied linguistics RAs. All these studies have evidenced the growth of citation frequency and the developing importance of citation in research writing.

2.2. Spread of citations in RA rhetorical sections

As pointed out in the previous section, citations in today's RAs are spread throughout the paper, although they still tend to cluster in the beginning section. RAs in the natural sciences and the social sciences are typically divided into four sections of distinct functions: Introduction, Methods, Results, and Discussion. The Introduction serves as a transition from the larger academic field to the particular study, and the Discussion directs the reader from the

particular study back to the larger academic area (Hill, Soppelsa, & West, 1982). It is, therefore, not surprising to find that references to previous work are much more abundant in these two sections than in Methods and Results. Thompson and Tribble's (2001) corpus study on agricultural botany doctoral theses produced a similar finding, with the Introduction and the Discussion containing noticeably more citations than the Methods and the Results sections.

Yet, as pointed out by Swales (1990) and evidenced in the RAs of many fields, the distinction of the two sections, Results and Discussion, can be a blurred one. Although many RAs have clear-cut sections labeled separately as Results and Discussion, some blend the two rhetorical sections into one. To complicate the matter even more, even in RAs with two clearly labeled sections of Results and Discussion, one can often find variation such as results simply redescribed in Discussion, and additional sections included and labeled as Conclusions, Implications, and so on. This diversity has somewhat complicated the research on citation use in RA rhetorical sections.

2.3. Surface forms of citations

Swales (1990) categorizes citations as integral and non-integral. Integral citations refer to the instances where the researcher's name appears as part of the sentence, either as a subject, a passive agent, part of a possessive noun phrase, or as "an adjunct of reporting" (Tadros, 1985, cited in Swales, 1990, p. 148). On the other hand, non-integral citations consist of references in which the researcher's name appears in parentheses or is represented by a superscript number. These two forms of citation can be used to show the degree of emphasis placed on a certain reference. For example, integral citations seem to give greater prominence to the cited author, while non-integral citations may imply an emphasis placed on the reported message (Hyland, 1999).

One factor affecting the choice between integral and non-integral citations, according to Swales (1990), is the difference in numerical or author/date citation conventions. He suggests that authors constrained by numerical conventions tend to adopt non-integral citations, a stance also taken up in Hyland (1999). However, Hyland's study of RAs in eight disciplines found that only philosophy papers employed more integral than non-integral citations, while the other seven disciplines invariably used more non-integral citations, showing less variation in the choice of citation forms across disciplines, although these disciplines apparently employ different citation conventions. Charles's (2006) study on theses in two contrasting disciplines, politics/international relations and materials science, found a similar lack of consistency between the choice of integral/non-integral forms and the use of citation conventions. Though the materials science theses adopting a numerical convention indeed used more non-integral than integral citations, out of seven politics theses adopting numerical conventions, six used far more integral than non-integral citations. The finding led Charles to conclude that the choice of integral/non-integral citation is complex and can be affected by citation convention, genre, discipline, and individual study type.

2.4. Cross-disciplinary and/or cross-cultural studies on citation

Another line of investigation that has received considerable attention is cross-disciplinary and/or cross-cultural studies (Adnan, 2004; Bloch & Chi, 1995; Hyland, 1999; Yao, 1998). Hyland's (1999) comparison of RAs in hard disciplines (biology, electronic engineering, physics, mechanical engineering) and soft disciplines (marketing, applied linguistics, sociology, philosophy) reveals a number of interesting findings. Compared to scholars in hard disciplines, writers in soft disciplines tended to use more citations. In addition, the latter group of academics was more likely to use integral citations, to place the author in subject position, and to employ direct quotes and discourse reporting verbs. Hyland attributed these

relatively more recursive citation practices to the different degrees of shared assumptions that hard and soft disciplines enjoy. Scientists in hard disciplines operate in "a relatively discrete and clearly identifiable area of study" (p. 352) and can thus "presuppose a certain amount of background, procedural expertise, theoretical understanding and technical lexis" (p. 353). With these shared assumptions, writers in hard disciplines do not rely so much on extensive reference to previous work for sustaining claims. On the other hand, researchers in the humanities and social sciences often have to "elaborat[e] a context through citation" (p. 353) in order to situate their research firmly within disciplinary frameworks and to sustain claims with "intertextual warrants" (p. 353).

Rather than exploring the hard-soft distinction (Hyland,1999), Adnan (2004) examined 30 Indonesian RAs in three humanities disciplines: linguistics, social and political science, and education. He found that Indonesian researchers used a low number of citations compared to those found in Hyland's (1999) study. The reasons for the low citation frequency were twofold. On the practical side, Indonesian academics have fewer resources and less time. On the cultural side, apart from a low interest in reading and writing in Indonesia, lacking an academic tradition also contributes to a lower number of citations in Indonesian research writing. Another interesting finding in Adnan's study was that education RAs used only about half as many citations as those in the other two investigated disciplines. His explanation was that all the education RAs compiled for the study concerned classroom teaching and problems in the teaching and learning processes. Since these were "practical" issues, it seemed legitimate, or at least acceptable to journal editors, for academics to ground arguments on "real world experience, rather than on a theoretical construct backed up with views and findings of experts in the disciplines" (p. 12).

2.5. Chinese scholars' citation behavior

Cross-cultural comparisons of RA writing have been made in a number of cultural contexts. This section will, however, focus on comparison studies of RAs written in Chinese or by Chinese scholars, and those written in English and/or by Anglo-American academics.

Comparing 31 RAs in the related fields of geophysics, metallurgy and mineral processing, materials science, and materials engineering, 11 in English from Euro-American mainstream journals, 10 in English by Chinese published in China, and 11 in Chinese by Chinese, Taylor and Chen (1991) found a notable difference in the number of references: with the Anglo-American English RAs employing markedly more references, and the China-published Chinese-language papers the fewest references. In their words, "the more 'Anglicized' the cultural milieu of the writing the greater emphasis on citing and summarizing the literature" (pp. 328–329). One of the proposed explanations was that Chinese scholars might be more resistant to citing and "exposing" shortcomings in others' work. Another simpler but perhaps more practical explanation was that Chinese scientists did not have much access to bibliographic information available in western countries.

Similarly, Bloch and Chi (1995) compared 120 articles (60 in English and 60 in Chinese) and identified three distinct functions of citations: giving background, giving support to an argument, and expressing a critical viewpoint. They found a consistently greater number of citations of all three functions in English-language texts than in Chinese-language texts. However, contrary to what had often been claimed, Chinese scholars did use critical citations in writing, although their critical position may be expressed in a more indirect way than in English articles. Bloch and Chi also investigated the age of the references used in the 120 RAs. Their underlying hypothesis was that the Chinese scientists, generally taught to value traditional texts, would use more older references. Results confirmed their hypothesis and showed that the English-language writers in the physical sciences used more recent references

than both the Chinese scholars in the physical sciences and the English-language writers in the social sciences. In other words, the citation date factor was found to be correlated with both language of writing and study area. Nevertheless, the cross-linguistic difference in the use of recent references was not explained merely as a cultural one, a point that will be explored further in a later section.

In Taiwan, Yao (1998) investigated the effect of disciplinary knowledge (humanities vs. life science) and language (Chinese vs. English) on the type, location, and function of citation use. RAs included in her corpus were limited to those published between 1991 and 1995. In the humanities corpus, 20 articles written in English were extracted from *TESOL Quarterly*, the flagship journal of the International TESOL Association; and 20 in Chinese from conference proceedings in the field of English teaching. It was found that Chinese RAs, on average, cited significantly fewer references than English RAs. In terms of rhetorical objectives, Chinese RAs employed fewer types of citations than their English counterparts. In addition to citation analysis, Yao (1998) conducted a survey to examine Taiwanese scholars' beliefs about citation use. An interesting finding of the survey study was that close to three-fourths of the respondents reported that they used more citations when publishing in English.

What these three studies have in common is their attempt at determining the impact of cultural variation and the language of writing on citation use. They reached a similar conclusion that Chinese researchers used fewer citations compared to their English-speaking counterparts, irrespective of disciplines. In addition, Chinese researchers used or claimed to use more citations when publishing in English, as compared with publishing in Chinese (Taylor & Chen, 1991; Yao, 1998). This finding suggests that the language of writing might be an influencing factor in citation use. It is therefore deemed necessary to further explore the

issue to understand Taiwanese scholars' citation behavior in English RA writing, as compared with those of their international counterparts.

3. Corpus and methodology

Two corpora were compiled for the current study. Texts included in the corpora were published between 2001 and 2005 and selected from two journals, *TESOL Quarterly* (TQ), one of the most prestigious international journals in the field of applied linguistics and language teaching; and *English Teaching and Learning* (ETL), a journal that enjoyed a favorable reputation in the field of English teaching in Taiwan. Several criteria were applied to text selection. First, Chinese RAs were excluded. Furthermore, only empirical studies were selected to be included in the two corpora. Theoretical and review articles were thus excluded. Finally, a total of 40 RAs, 4 from each journal each year, were prepared for this corpus study. All the abstracts, appendices, references, tables, and figures were removed so as not to distort results. A special note should be made concerning the labels of "Taiwanese scholars" and "international scholars." In this study "Taiwanese scholars" and "international scholars" were defined arbitrarily according to the affiliations of the authors. The journal *English Teaching and Learning* was based in Taiwan, with contributions predominantly from Taiwanese researchers. Accordingly, the 20 RAs included in the ETL corpus were all contributed by Taiwanese researchers. On the other hand, *TESOL Quarterly* authors were more varied in terms of the countries of affiliations. The 20 RAs included in the TQ corpus therefore covered contributions from researchers based in the United States, the United Kingdom, China, Hong Kong, Singapore, Canada, and New Zealand. Table 1 shows the details of the two corpora.

Table 1. Details of the Corpora

Corpus	Number of texts	Average length of texts (in words)	Total size of corpus (in words)
TQ	20	7,289	145,777
ETL	20	5,342	106,848

Citations in each rhetorical section—Introduction, Methods, Results, and Discussion—were identified and recorded. They were further categorized as either integral or non-integral. In addition, each citation was recorded, according to its presentation, as quote, block quote, summary/paraphrase, or generalization. Finally, the date of each reference was examined to determine its recency. Frequency analysis was performed and comparison was made between the two corpora in order to determine the similarities and differences between Taiwanese and international scholars' citation behavior.

In this paper, a "citation" is defined as "a research report which has a specific reference point that is clearly identifiable" (Charles, 2006, p. 314). Furthermore, each occurrence of a research report counts as a separate citation. Therefore, in the following example, the repeated occurrences of the same reference were counted as two citations.

(1) Davies (1991) argues that membership as a NS in a speech community is a matter of self-ascription. Instead of labeling speakers from an analyst-driven perspective, Davies (1991) suggests defining NSs as people who have a "special control over a language, insider knowledge about 'their' language" (p. 1). (TQ12)

4. Results and Discussion

Table 2 reveals that, on average, TQ RAs use more citations than ETL RAs (10.6 vs. 8.3 per 1,000 words). It is also found that differences exist in the number of citations used by individual RAs in the same corpus. For example, TQ1 includes as many as 137 citations, 3.6

times the citation number used in TQ8. On the other hand, in the ETL corpus, ETL18 employs 19 citations, a quarter of the citation number in ETL7. This indicates that although TQ articles generally use more citations, individual TQ articles may include fewer citations than RAs in the ETL corpus.

Table 2. Number of Citations

Corpus	Average per paper	Range	Total	Citations per 1,000 words
TQ	77.3	38–137	1,546	10.6
ETL	44.5	19–76	891	8.3

4.1. Surface forms of citations

Hyland's (1999) study points out an overall tendency to use non-integral citation forms in RAs across disciplines. For example, applied linguistics papers in his corpus used twice as many non-integral as integral citations (65.6% vs. 34.4%). In the current study a similar pattern was found. Non-integral forms comprise 66.9% of the total citations in the TQ corpus, doubling the number of integral forms (Table 3). The ETL papers use slightly fewer non-integral citations than the TQ RAs (60.3% vs. 66.9%), but there are still significantly more non-integral than integral forms in the ETL corpus.

Table 3. Surface Forms of Citations

Corpus	Non-integral	Integral
TQ	1035 (66.9%)	511 (33.1%)
ETL	537 (60.3%)	354 (39.7%)

4.2. Presentation of cited work

The study then analyzed how citations are incorporated in the RAs in the two corpora. The cited work can be presented as either quotes (within-sentence quotations), block quotes (block

and indented quotations of more than 40 words), paraphrase/summary, and generalization (combining several sources) (Swales & Feak, 2000). Table 4 shows how citations are incorporated into the articles in the two corpora.

Table 4. Presentation of Cited Work

Corpus	Quote	Block quote	Summary / Paraphrase	Generalization
TQ	146 (9.4%)	21 (1.4%)	636 (41.1%)	743 (48.1%)
ETL	31 (3.5%)	1 (0.1%)	450 (50.5%)	409 (45.9%)

Analysis reveals that citations are overwhelmingly expressed as summary and generalization (89.2% and 96.4% in TQ and ETL corpora respectively). This finding corroborates with Hyland's (1999), where a comparison of RAs in eight disciplines shows that citations were chiefly framed as summary and generalization in the articles investigated, apparently because these two forms "allow the writers greater flexibility" (p. 26) to comment on the message they are citing. The following two extracts, drawn from the most densely cited RA in the two corpora, respectively, demonstrate such use of summary and generalization. Extract 2 employs one generalization citation and one summary citation, while Extract 3 includes two summary citations:

(2) Similarly, a more abstract level of multicultural knowledge should include an understanding of how knowledge itself is created, how it can be viewed as a social construction, and how it can be a product both of a unique culture and of a particular power structure (Banks, 1991; Dana, 1993). As Davidman (1990) suggests, knowing this about knowledge sets the stage for teachers to be critical of their perspective on society. Teachers who have majored in the social sciences and teach in these areas should be more likely than others to develop this abstract understanding of knowledge. (TQ1)

(3) Lee (1998) points out the possible ways of collecting formative learning information, which include classroom observation, tasks, open-ended questions, learning portfolios, checklists and paper-and-pencil tests. Popham (1995) further asserts that students' real learning can only be exhibited through multiple evidence. (ETL9)

Although both corpora were abundant in summary and generalization citations, their use of direct quotes and block quotes exhibits a different pattern. While 10.8% of the citations in the current TQ corpus involve a quotation, the ETL corpus has only 3.6% of citations in the form of quotations. The investigation found only one paper in the TQ corpus that handles all the citations in summary and generalization. The others employ either quotes or block quotes or both to support their argument, among which TQ2 uses as many as 21 short quotes and TQ9 five block quotes to present the cited work. On the other hand, 9 out of 20 RAs in the ETL corpus do not include any short or extended quotes, and only one instance of block quote was found in the entire corpus. The following extract from the TQ corpus shows that in a stretch of 73 words the author incorporates three short quotes, totaling 30 words.

(4) Pennycook's *The Cultural Politics of English as an International Language* (1994) examined further issues of power and dominance with regard to English, questioning the view of its spread as "natural, neutral, and beneficial" (p. 6) and emphasizing that English is "bound up in a wealth of local social, cultural, economic and political complexities" (p. 7). Pennycook also discussed "aspects of resistance and human agency in appropriating English to its local contexts" (p. 7). (TQ2)

How do we account for this marked difference in the use frequency of short and block quotes between the two corpora? One explanation is that it may reflect personal preference and writing style. However, L2 researchers have been documented to experience difficulty in expressing ideas due to a less rich vocabulary and consequently they deliberately keep to a

simple style (Flowerdew, 1999). In a similar vein, the Taiwanese researchers may feel that it is more difficult to integrate others' words successfully, thereby opting to summarize or paraphrase and take advantage of the flexibility allowed in these two forms of presentation. Nevertheless, as this study involves only textual analysis, authors' motivation for these various citational forms can only be guessed at.

4.3. Citations in RA rhetorical sections

In analyzing the spread of citations in rhetorical sections, I first referred to the IMRD (Introduction-Methods-Results-Discussion) macrostructure (Hill, Soppelsa, & West, 1982) that has become well established in the world of research writing. However, the initial attempt to identify rhetorical sections encountered some complications. A total of eleven RAs (one in the TQ and ten in the ETL corpus) were found to contain a blended Results and Discussion section, which made it difficult to compare between the two corpora. Since this distinctive Results and Discussion section has been explored and found to serve primarily similar functions as the conventional Results section—report and comment on research findings—and since the conventional Discussion section chiefly comments on results (Yang & Allison, 2003), it was decided that citations occurring in the sections following Methods, including Results, Discussion, Results and Discussion, Conclusion, Pedagogical Implications, and others, were counted together under the merged category Results and Discussion.

As revealed in Table 5, the pattern of citation use between the two corpora is similar. The difference in the percentage of citations occurring in Methods and Results and Discussion is rather insignificant. On the other hand, around two-thirds of citations are found in Introduction sections in both TQ and ETL RAs. Since most obligatory moves in the RA Introduction involve reviewing and indicating a gap in previous research (Swales & Feak, 2000), it is hardly surprising to find an abundance of citations in this section. On the other

hand, RA authors tend to compare findings with previous work (Yang & Allison, 2003), thereby incorporating more citations in Results and Discussion than in Methods. Notwithstanding this general tendency, a scrutiny of individual RAs reveals some marked differences, even among RAs in the same corpus. While most RAs follow the general pattern, some adopt rather unconventional citational use. For example, in TQ8, an experimental study, 37 out of 38 citations appear in the Introduction section, while 81.9% of all citations in TQ11, a study of English educational policies, are found in Results and Discussion. In an ETL corpus, ETL5, a survey study, cites previous research copiously (46.7% of all citations) to compare results and to lend support to pedagogical suggestions. In the same corpus, however, four out of twenty RAs do not include any citations in Results and Discussion.

Table 5. Citations in RA Rhetorical Sections

Corpus	Introduction	Methods	Results & Discussion
TQ	1015 (65.6%)	191 (12.4%)	340 (22%)
ETL	582 (65.3%)	86 (9.7%)	223 (25%)

4.4. Citation dates

Citation dates refer to the time difference between the publication of the citing work and that of the cited reference. This difference may indicate the degree of the value placed on more traditional texts or more recently published articles. In scientific articles, most references were found to fall within five years (Price, 1986, cited in Bloch & Chi, 1995). Bloch and Chi's (1995) comparison study found that English articles used more recent citations than their Chinese counterparts and that papers in the physical sciences contained newer references than those in social sciences. While their findings point to the language and discipline factors in the recency of the cited references, the investigation of the citation dates in the current study also reveals an interesting discrepancy between the two corpora of the same language and the

same discipline area. As Table 6 shows, the two corpora contain a similar number of citations aged between five and fifteen years. However, the TQ papers use more recent citations than the ETL counterparts (32% vs. 22.9%). Conversely, the ETL articles employ a greater number of older references than those in the TQ corpus (27.4% vs. 19%). In other words, international scholars tend to include more updated information and avoid ancient citations than their Taiwanese counterparts.

Table 6. Citation Dates

Corpus	5 years or less (%)	5 to 15 years (%)	15 years or older (%)
TQ	32	48.9	19
ETL	22.9	49.6	27.4

Yet, again, in this aspect of investigation, individual variance can easily be detected both in the TQ and ETL corpora. In the ETL corpus, ETL9 cites 22 recent references, slightly over half of its full reference list (51.2%). On the other hand, ETL19 contains not a single recent reference, while ETL4 and ETL18 have one each. On the TQ side, a similar divergence can be identified. While TQ9 employs as many as 40 recent references (57.1%), TQ8 cites only two newer studies (5.7%). Still, overall, the TQ researchers cite more newer references than their ETL counterparts.

While accounting for the more dated citations in Chinese language articles than in English language ones, Bloch and Chi (1995) offered two explanations. The first was that Chinese writers were more "tied to the past" and therefore more likely to "rely on older, classic texts" (p. 248). The second explanation Bloch and Chi advanced was that Chinese writers, particularly at the time of their investigation, had less access to more updated research studies. However, in the current study neither of the arguments can account for the finding satisfactorily. In Taiwan, applied linguists, studying a subject originating in the West, rely mostly on foreign publications for theoretical frameworks and research methodologies.

Granted, foreign publications are generally more difficult to locate than those published in the country. However, the current research environment has largely made it possible for scholars to obtain publications through purchases or library loans. Particularly in the Information Age, the burgeoning of electronic publication, including journals available in electronic form, has increased the possibility that scholars based in Taiwan can also keep pace with advancements in the international academic community. Therefore, one explanation to account for the fewer recent citations in the ETL corpus can be that Taiwanese may not always aim to keep ahead of the fast-moving progress of information or include the most recent research updates in their papers.

5. Conclusion

This corpus-based study used 40 RAs selected from two TESL journals, *TESOL Quarterly* and *English Teaching & Learning*, to explore and compare international and Taiwanese scholars' citation use. The findings corroborated previous studies that Chinese-speaking scholars cite less than their English-speaking counterparts (Bloch & Chi, 1995; Taylor & Chen, 1991; Yao, 1998). However, the results also show that in several aspects investigated, the two corpora generally employ similar strategies in incorporating the cited information. For example, both corpora use significantly more non-integral than integral citations. They both present around 90% of the cited information in the form of summary/paraphrase and generalization. Around two thirds of citations are found in Introduction sections in both TQ and ETL corpora. This suggests that scholars working in the same disciplines, notwithstanding their dissimilar cultural or national background, are more likely to abide by a similar set of conventions when constructing knowledge and study reports (Bloch & Chi, 1995; Hyland, 1999; Swales & Lindemann, 2002; Taylor & Chen, 1991). Furthermore, Taiwanese applied linguists are mostly educated in the West, read publications in English,

perhaps also observe and imitate how international scholars write, thereby following similar discourse conventions in how they cite. In a discussion of "international scientific English", Wood (2001) suggests a similar viewpoint that, before writing their first English papers, experienced second-language writers would have become familiar with the conventions of their field through reading numerous papers in English. How to use citation properly and strategically should be one of such conventions learned either explicitly or implicitly.

Alongside some similarities, the study also reveals a number of differences between the two corpora. Overall, Taiwanese scholars use fewer citations than their international peers. Certainly the density of citation may not bear a direct relation with the quality of the RA; in fact, too many references, particularly when they are handled uncritically, can sometimes obscure the citer's original contributions (Belcher, 1994). Yet, as an important function of citation is to situate one's own study in the research tradition and to persuade the reader of the soundness and novelty of the citer's arguments (Hyland, 1999), a failure to cite significant works may not reflect well on the value of the study. Another difference between the Taiwanese and international scholars' citation behavior involves their use of direct quotes. The results suggest that Taiwanese applied linguists include fewer short and block quotes in their RAs, as compared with their international peers. Although it is speculated that Taiwanese writers may prefer to summarize or paraphrase the cited information in order to maintain their own writing style, the motivation behind this rhetorical choice needs further exploration. A third disparity in the two corpora is the recency of the references. The analysis found fewer recent and more dated references in the ETL corpus. As not "up-to-date" sources can be considered a problem (Swales & Feak, 2000) and may indicate the researcher's lack of awareness of the most recent development in the research issue, Taiwanese applied linguists may need to seek improvement in this aspect if they wish to publish in international journals.

While a general tendency was found in this corpus study, the investigation also noted individual differences among RAs in the same corpus. For example, while TQ papers overall employ more citations than ETL articles, individual ETL papers may include twice as many citations as individual TQ articles. While 20% of the ETL RAs do not refer to another's work in Results & Discussion, a number cite copiously to compare their own findings with those in the previous studies. In the age of the references cited, of all the references cited in ETL9, over half of them fall within five years of the others; yet another RA of the same corpus cites not a single recent reference. Similar individual variance has also been found in Hemais (2001) and Thompson and Tribble (2001). Whether this variance is due to personal writing style or the nature of the research (Thompson & Tribble, 2001) is not within the scope of the current study. Further research with a wider sample of articles or with an ethnographic approach may contribute to a more thorough understanding.

As this study utilized only a limited number of RAs selected from two TESL journals, the results need to be interpreted with caution. The present findings have revealed some congruence and discrepancies between the two investigated corpora, but they also point to several areas of curiosity that can be further explored. The first concerns the language factor. Do the same researchers exhibit the same citation behavior when writing in different languages? Furthermore, although the literature has shown disciplinary differences in citation practices (e.g., Hyland, 1999), do Taiwanese RAs across disciplines vary in a similar way when referring to another's work? Finally, although the current study has revealed individual differences, it cannot determine the motivation behind the differing practices. A survey study or an in-depth interview study may help identify writers' motivation in including and selecting various citational forms.

References

Adnan, Z. (2004). Citing behaviours in Indonesian humanities research articles. *ASAA E-journal of Asian Linguistics & Language Teaching, 6*. Retrieved May 14, 2006, from www.arts.unsw.edu.au/languages/asaa-ejournal/Issue6/Adnan.pdf.

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison: University of Wisconsin Press.

Belcher, D. (1994). The apprenticeship approach to advanced academic literacy: Graduate students and their mentors. *English for Specific Purposes, 13*(1), 23–34.

Bloch, J., & Chi, L. (1995). A comparison of the use of citations in Chinese and English academic discourse. In D. Belcher & G. Braine (Eds.). *Academic writing in a second language: Essays on research and pedagogy* (pp. 231–274). Norwood, NJ: Ablex.

Charles, M. (2003). ”This mystery”: A corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes, 2*, 313–326.

Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes, 25*(3), 310–331.

Cronin, B. (1998). Metatheorizing citation. *Scientometrics, 43*(1), 45–55.

Flowerdew, J. (1999). Problems in writing for scholarly publication in English: The case of Hong Kong. *Journal of Second Language Writing, 8*(3), 243–264.

Gilbert, G. N. (1977). Referencing as persuasion. *Social Studies of Science, 7*(1), 113–122.

Harwood, N. (2005). *"Nowhere has anyone attempted ... In this article I aim to do just that"*: A corpus-based study of self-promotional *I* and *we* in academic writing across four disciplines. *Journal of Pragmatics, 37*(8), 1207–1231.

Hawes, T., & Thomas, S. (1997). Tense choices in citations. *Research in the Teaching of English, 31*(3), 393–414.

Hemais, B. (2001). The discourse of research and practice in marketing journals. *English for Specific Purposes, 20*, 39–59.

Hill, S., Soppelsa, B., & West, G. (1982). Teaching ESL students to read and write experimental research papers. *TESOL Quarterly, 16*, 333–347.

Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics, 20*(3), 341–367.

Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes, 20*(3), 207–226.

Kuo, C. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes, 18*(2), 121–138.

Leydesdorff, L. (1998). Theories of citation? *Scientometrics, 43*(1), 5–25.

Myers, G. (1990). *Writing biology: Texts in the social construction of scientific knowledge*. Madison: University of Wisconsin Press.

Pennycook, A. (1994). *The cultural politics of English as an international language*. Harlow: Longman.

Salager-Meyer, F. (1994). Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes, 13*(2), 149–170.

Salager-Meyer, F. (1999). Referential behavior in scientific writing: A diachronic study (1810–1995). *English for Specific Purposes, 18*(3), 279-305.

Shaw, P. (1992). Reasons for the correlation of voice, tense, and sentence function in reporting verbs. *Applied Linguistics, 13*(3), 302–319.

Swales, J. (1990). *Genre analysis: English in academic and research settings.* Cambridge, UK: Cambridge University Press.

Swales, J. M., & Feak, C. B. (2000). *English in today's research world*. Ann Arbor: University of Michigan Press.

Swales, J. M., & Lindemann, S. (2002). Teaching the literature review to international graduate students. In A. Johns (Ed.), *Genre in the classroom: Multiple perspectives* (pp. 105–119). Mahwah, NJ: Lawrence Erlbaum.

Taylor, G., & Chen, T. (1991). Linguistic, cultural, and sub-cultural issues in contrastive discourse analysis: Anglo-American and Chinese scientific texts. *Applied Linguistics, 12*(3), 319–336.

Thompson, G.., & Ye, Y. (1991). Evaluation in the reporting verbs used in academic papers. *Applied Linguistics, 12*(4), 365–382.

Thompson, P., & Tribble, C. (2001). Looking at citations: Using corpora in English for academic purposes. *Language Learning & Technology, 5*(3), 91–105.

White, H. D. (2004). Citation analysis and discourse analysis revisited. *Applied Linguistics, 25*(1), 89–116.

Wood, A. (2001). International scientific English: The language of research scientists around the world. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 71–83). Cambridge, UK: Cambridge University Press.

Yang, R., & Allison, D. (2003). Research articles in applied linguistics: Moving from results to conclusions. *English for Specific Purposes, 22*, 365–385.

Yao, C-K. (1998). Citation use in research articles: A comparison of English and Chinese research articles in the fields of humanities and life science. NSC Research Project Report (NSC88-2411-H-034-009).

Acknowledgements

Research for this article was sponsored by grants from the National Science Council, Taiwan (NSC95-2411-H-194-020).

The Asian ESP Journal

Authenticity in ESP/EAP Selection Tests Administered at Iranian Universities

Goudarz Alibakhshi, *Yasouj University*

Gholam Reza Kiani, *Tarbiat Modares University*

Ramin Akbari, *Tarbiat Modares University*

Biodata

Goudarz Alibakhshi is an Assisstant Professor of applied linguistics at Yasouj University, Yasouj, Iran. His areas of interests are ESP and language assessment.

Gholam Reza Kiani is an Associate Professor of Applied Linguistics at Tarbiat Modares University, Tehran, Iran. He has been teaching ESP, language testing and research methodology since 15 years ago.

Ramin Akbari is an Assistant Professor of Applied Linguistics at Tarbiat Modares University, Tehran, Iran. His main interest areas are teacher education and Applied Linguistics.

Abstract

The issue of authenticity has always been an important part of any discussion on language testing. Authenticity, situational and interactional, is also a critical quality of English for Specific/Academic Purposes (ESP/EAP). Despite the significance of this quality, only a few studies have been carried out to investigate ESP/EAP tests in terms of authenticity. The present research was an attempt to study the authenticity characteristics of ESP/EAP tests administered as a part of MA/MSc or PhD entrance examination in 2006 and 2007 at Iranian

universities. A mixed design study, both qualitative and quantitative, was used. The data were gathered through a semi-structured interview and a questionnaire. In addition, ESP test tasks were compared with characteristics of target language use situations using the framework provided by Bachman and Palmer (1996). The results were analyzed through descriptive and inferential statistics. The results indicated that ESP/EAP tests administered at Iranian universities are not authentic. That is, they are not representative of target language use situation tasks and the learners' language needs. The participants also indicated that a change in the purpose of ESP tests is needed.

Keywords: ESP/EAP tests; authenticity; test tasks; and learners' needs

1. Introduction

The notion of authenticity has been much discussed within the fields of applied linguistics and general education. In applied linguistics this concept emerged in the late 1970s, when communicative methodology was gaining significance and there was a growing interest in teaching and testing 'real-life' language (Lewkowicz, 2000). Bachman (1991) makes a distinction between situational and interactional authenticity. Situational authenticity is derived from an analysis of tasks in the target language use situations, the features of which are realized as test task characteristics. Therefore, situational authenticity can be demonstrated by making explicit the relationship between the test task characteristics and the features of the tasks in the target language use (TLU) situation. On the other hand, interactional authenticity is a function of the extent and type of involvement of test-takers' ability to accomplish a test task. Bachman and Palmer (1996) and Douglas (2000) consider authenticity a critical quality of language tests alongside validity, reliability, consequences (impact and washback), interactivenes, and practicality.

Authenticity is, therefore, considered to be an important test quality because it relates the test task to the domain to which we want our score interpretations to generalize. Authenticity thus provides a means for investigating the extent to which score interpretations generalize beyond performance on the test to language use to the TLU domain, or to other similar non-test language use domains. Bachman and Palmer (1996) suggest that "This links authenticity to construct validity, since investigating the generalizability of score interpretations is an important part of construct validation" (p. 24).

In the ESP testing situation, authenticity of task refers to the similarity of the test tasks to the target language use situation. The more authentic a text or a task is, the more probable it is for the test-taker to carry out the test tasks in the same way, as he/she would perform it in the actual target situation. Thus, authentic texts and tasks are representative of the specific language use situation or similar to the tasks a test-taker actually performs in his/her work. Authenticity is also of high significance to Douglas's (1997) consideration of specific purpose tests in that it is one of the two features which distinguish such tests from more general purpose tests of language (the other feature being the interaction between language knowledge and specific purpose content knowledge).

However, despite the importance accorded to authenticity, there has been a marked absence of research to demonstrate this characteristic, particularly in ESP tests. Therefore, the present study is an attempt to study the authenticity of ESP tests used at universities in Iran to measure Iranian post-graduate students' competence in ESP/EAP.

1.1 Objectives of the study

The study examines the ESP tests administered as a part of MA/MSc and PhD entrance examination in 2006 and 2007 in terms of authenticity characteristics. The MA/MSc tests were administered nationally at all state universities; the PhD examinations were administered

at the three state universities of Tarbiat Modares, Tehran, and Shahid Chamran. Simply put, the main objectives of the study were to provide a systematic analysis of these tests in terms of setting, test rubrics, test input, outcomes, and the relationship between input and response. Besides, the study also aimed to compare the tasks of the tests with the learners' language needs, that is, to determine the authenticity of test tasks. It also invited test-takers to comment on the authenticity of these tests. In doing so, the following research questions were raised:

1. To what extent the contents and tasks of the current EAP tests conducted in graduate and post-graduate programs are authentically representative of the tasks in target language situations?
2. To what extent are the skills measured by the ESP tests representative of the students' language needs?
3. What is the perception of the students and teachers of the authenticity of the ESP tests?

2. Review of literature

The issue of authenticity has always been an important part of any discussion on language testing. There is an argument that only "real language use" should be counted when assessing language ability. 'Authenticity', in early communicative texts, meant the use of material in prompts that had not been written for non-native speakers of English, and a test could not be communicative unless it was authentic (Fulcher, 2000). Alderson (1981) regards authenticity a "sterile argument"(p. 48). Modern performance tests that attempt to mirror some criterion situation in the external world are no more than role-plays or simulations, in which the learner is asked to 'imagine' that they are actually taking a patient's details in a hospital, giving students a mini-lecture, or engaging in a business negotiation. Language tests, by their very

nature, are not mirrors of real life, but instruments constructed on the basis of a theory of the nature of language, language use, and language learning (Fulcher, 2000).

Widdowson (1978), making a distinction between 'genuineness' and 'authenticity' of language testing, argues that "genuineness is a characteristic of the passage itself and is an absolute quality. Authenticity is a characteristic of the relationship between the passage and the reader and has to do with appropriate response" (p 80). Widdowson (1990; 1994; 1998) also maintains that genuineness is viewed as a quality of all texts, while authenticity is a characteristic 'bestowed' on texts by a given audience. In his view, authenticity is a quality of a text if the audience could comprehend the author's intentions, which would only be possible when the audience is aware of the conventions used by the writer or speaker.

Lynch (1982) argues that despite the variety of interpretations of authenticity and authentic assessment, one feature of authenticity upon which there has been general congruence over time is that it is a significant quality for test development which 'carries a positive charge'.

Throughout the 1980s the authenticity debate laid more emphasis on the nature of test input with scant regard being paid to the role test-takers play in processing such input. The debate centered on desired qualities of those aspects of language tests which test setters control, with advocated authenticity as promulgating the use of texts and tasks taken from real-life situations (Morrow, 1983, 1991; Carroll, 1980; Doye, 1991). The skeptics draw attention to the limitations of using such input and the drawbacks associated with equating such input with real-life language use (Alderson, 1981; Davies, 1988, 1984; Spolsky, 1985).

In language teaching the debate is taken forward by Breen (1985) who suggested that authenticity may not be a single unitary notion, but one relating to texts (as well as to learners' interpretation of those texts), tasks and social situations of the language classroom. Breen draws attention to the fact that the aim of language learning is to acquire the ability to

interpret the meaning of texts, and that any text which moves towards achieving that goal could have a role in teaching. He maintains that the notion of authenticity was a fairly complex one and that it is over-simplistic to dichotomize authentic and inauthentic materials, particularly since authenticity was, in his opinion, a relative rather than an absolute quality.

Bachman (1991) reaffirms Widdowson's notion of authenticity as a quality of outcome arising from the processing of input, but at the same time he points out a need to account for 'language use' which Widdowson's unitary definition of genuineness does not address.

Bachman (1990; 1991) also recognizes the complexities of authenticity, arguing that neither situational nor interactional authenticity is absolute. A test task could be situationally highly authentic, but interactionally low on authenticity, or vice versa. This reconceptualization of authenticity into a complex notion pertaining to test input as well as the nature and quality of test outcomes is similar to the view of authenticity emerging in the field of general education. In the United States, in particular, the late 1980s/early 1990s saw a movement away from standardized multiple-choice tests to more performance-based assessment characterized by assessment tasks which are holistic, provide an intellectual challenge, are interesting for the students, and are tasks from which students could learn (Carlson, 1991).

The authenticity debate seems to have moved in a somewhat different direction in 1996. In language education, Bachman and Palmer (1996) argue that the notion of authenticity is different from interactiveness. They define authenticity as "the degree of correspondence of the characteristics of a given language test task to the features of a TLU task" (p. 23). This definition corresponds to that of situational authenticity, while interactiveness replaced what is previously called interactional authenticity. The assumption behind this direction is an understanding that all real-life tasks are by definition situationally

authentic, so authenticity can only be an attribute of other tasks, that is, those used for testing or teaching. At the same time, not all genuine language tasks are equally interactive; some give rise to very little learning of language. However, authenticity is in part dependent on the correspondence between the interaction arising from test and TLU tasks. Perceiving the two as separate entities may, therefore, be misleading. Douglas (2000) continues to see the two as different aspects of authenticity, arguing that both need to be present in language tests for specific purposes.

To approximate the degree of correspondence between tests and TLU tasks, that is, to determine the authenticity of test tasks, Bachman and Palmer (1996) propose a framework of task characteristics. The framework is important since it provides a useful checklist of task characteristics, which allows for a degree of agreement among test developers interested in ascertaining the authenticity of test tasks. It takes into account both the input provided in a test as well as the expected outcomes arising from the input by characterizing not only test tasks but also test-takers' interactions with these tasks.

3. Method

3.1. Participants

Three groups of 435 participants took part in the study. The first two groups were 150 doctoral and 240 master students of chemistry, biology, geography, civil engineering, and Persian literature and geography. The third group was made up of 45 subject specialists in the aforementioned fields of study. The students were selected from Tarbiat Modares, Tehran, Allama, Shiraz, Shahid Chamran, Tarbiat Moalem and Shahid Beheshti Universities. The subject specialists were invited to participate because they are responsible for teaching ESP at these universities. All the participants were aware of the purpose of the study and were willing to participate.

3.2. Tests

The tests selected to be analyzed were the ESP tests administered at all state universities to select master candidates of chemistry, geography, biology, Persian literature, and civil engineering. These tests were developed and administered by the Assessment Organization of Iran in 2007 and 2008. In addition, different ESP tests developed and administered to select the PhD students of the aforementioned disciples at Shahid Chamran, Tarbiat Modares, and Tehran Universities were also used. For each of the disciplines, two or three ESP tests administered and one component of the PhD test was selected, reviewed, and analyzed. ESP tests for MA/MSc candidates include technical reading passages, followed by multiple-choice items. ESP tests for PhD students include reading passages which the test-takers were to translate into Persian. Some of them include native to target language translation tasks. (Refer to the appendix for sample tests).

3.3 Procedure

This was a qualitative/quantitative study. In order to conduct the quantitative part of the study, i.e, to determine the correspondence between target language use situation tasks and ESP/EAP test tasks, the framework adapted from Bachman and Palmer (1996) and Douglas (2000) was used. This framework provides a systematic way of matching tasks in terms of their setting, test rubrics, test input, the outcomes the tasks are expected to give rise to, and the relationship between input and response. Moreover, in order to see whether the ESP tests are representative of the learners' language academic needs, the skills measured by such tests were compared with the learners' academic needs analyzed and assessed by different scholars, especially Farhadi (2008).

In addition, 240 MA/MSc, 150 PhD students, and 45 instructors were invited to comment on the authenticity of these tests as well as the purposes which the ESP tests serve. Two main instruments were employed to collect data of this part of study: an interview and a questionnaire. The questionnaire used in this study was a component of a washback questionnaire developed and validated by Alibakkshi, Kiani, & Akbari (2010). It was designed on a four-point Likert scale which invited the participants to comment on their perceptions of authenticity of ESP tests. Interview data were obtained through structured, individual interviews in Iran at the universities mentioned before with 10 subject specialists and semi-structured individual interviews with the 30 MA/MSc and 15 PhD students. Altogether 55 interviews were conducted and transcribed. The quantitative data of the study were analyzed through descriptive and inferential statistics (Chi-square test). The transcribed interviews were content analyzed and the main themes were extracted.

4. Results of the study

4.1 Quantitative results

4.1.1 Task Authenticity

To determine whether there is a significant correspondence between ESP tests tasks and target language use situation tasks, the ESP tests were carefully analyzed. As the ESP tests administered to select different MA/MSc and PhD candidates of the aforementioned universities were all the same in terms of characteristics of test tasks and number of items (the sample tests for all the majors are put in Appendix 1), only the test for students of chemistry was analyzed in this part of the study, and the results were generalized to the other tests. The comparison between the target language situation tasks of chemistry students and the chemistry test tasks is represented in Table 1.

Table 1: Comparison of task characteristics of the TLU situation and an ESP test for chemistry students

Characteristics	Target language use situation tasks	Test tasks
Rubrics Objectives	To review homework problems, teach students how to do experiments, explain the contents of lessons, and receive feedback from learners.	To check test-takers' reading comprehension of technical content
Procedures for responding	Interact orally and on whiteboard with students to explain new concepts and complete experiments in the lab	Reading passages and selecting the best choices
Structure Number of tasks Task distinction	Varies, problems raised by students and teachers, introducing new concepts, and doing different types of experiments Students' problems and teachers' comments may come at any time. Tasks may blend into each other	Number of tasks is predictable. There are some reading passages and the test-takers should answer the questions related to each passage. The tasks are all the same.
Characteristics of Input Features of context setting	Typical chemistry classroom: instructor behind teacher table, blackboard, overhead projector; students by desks, have notebooks and textbooks or in a typical chemistry lab; instructor moving to each student; lab equipment, etc.	A typical testing setting, the test-takers sitting on chairs, they all are of the same major
Participants	Instructors, all chemistry students, lab assistants	Test-takers and some proctors
Tone	Sometimes informal and friendly; other times formal	Always formal language
Format	Oral, sometimes written material in textbook, lecture notes, oral questions from teachers and students	Written
Vehicle of delivery	Live and written	Written and not live
Length	90-minute class, a long discussion between teachers and students, lectures, or reading specific textbooks, etc.	A 30-minute testing session (master exam) and a 60-minute testing session (PhD exams)
Characteristics of expected response Format	Oral explanation of problems, writing on blackboard, taking notes, writing the summary of textbooks, doing experiments following commands	Written and recognition (in master tests) and translation (in PhD exams)
Channel	Oral and written	Written
Form	Language and sometimes non-language (gestures, actions, etc.)	Only language
Language length	Varies, single words, phrases, paragraphs, and extended discourse	Single words and limited discourse in master tests; limited discourse in PhD tests
Type	Limited and extended production	Limited recognition in master exams but limited production in PhD exams
Reactivity	Highly reciprocal when the channel is oral	Nonreciprocal
Scope	Very broad: high degree of input must be processed	Very narrow: small degree of input must be processed
Directedness	Sometimes direct and sometimes indirect	Indirect

Having carefully analyzed chemistry test tasks and target language use situation tasks of these students, we argue that in terms of all the test characteristics above, such as rubrics, input, output, etc., there is little correspondence between test tasks and target language use situation tasks. Such interpretation could, due to close similarities between ESP tests for different fields of study in terms of test task types, be generalized to all ESP selection and placement tests administered in Iran. That is, the ESP tests developed for the students of Persian literature, civil engineering, biology, and geography are not representative of target specific/academic use situation tasks. For instance, the vehicle of delivery in target language use situation is live and written while that of ESP tests is only written, and the expected response of TLU situation tasks is highly reciprocal while that of ESP test tasks is nonreciprocal (see the above table for more details). The interviews with two applied linguists and analysis in eight content fields also verified that these so-called ESP tests are not authentically representative of target specific/academic language use situation tasks.

4.1.2 Correspondence between learners' language needs and the domain of ESP tests

Another major issue in authenticity of ESP tests is the correspondence between learners' language and educational needs and the contents of ESP tests. The checklist of master and doctoral students' language needs adapted from different need analysis studies, particularly Farhadi (2008), and the language skills and sub-skills measured by ESP tests are represented in the following table.

Table 2: Comparison of learners' language needs and the domain of ESP tests

Language and academic needs of the learners	**The skills measured by the tests**
General and technical listening, reading, speaking, and writing, note taking, internet search, reading graphs and tables, summarizing, translation, writing papers, submitting proposals, presenting papers in conferences, writing reports, letter writing, and writing CV	The skills measured by ESP tests administered in Iran are technical reading in tests developed to select master candidates and translating in PhD tests

As the results in Table 2 indicate, there is no correspondence between the skills measured by ESP tests and language needs of Iranian master and doctoral students of the aforementioned fields of study. In other words, it could be said that the main language needs of language learners are: general and technical listening, reading, speaking, and writing, note taking, internet search, reading graphs and tables, summarizing, translation, writing paper, submitting proposals, presenting papers in conferences, writing reports, letter writing, and writing CVs; whereas, the only skill measured by ESP tests administered in Iran is technical reading in tests developed to select master candidate and translation in PhD tests.

4.1.3 Teachers and students' perceptions of the authenticity of the ESP tests

In order to invite the participants to comment on the authenticity of the ESP tests, the following questions, designed on a four-point Likert scale of very low, low, average, and high, were asked. For the purpose of analysis, the two scales of very low and low were converted into one scale, low.

1. To what extent are the ESP tests administered locally representative of the students' language needs in target situations?
2. To what extent do the test tasks in ESP tests correspond to the target language use situation tasks?
3. To what extent are the language skills measured by ESP tests representative of the language skills and sub-kills needed by the learners to do their own academic tasks?

The descriptive statistics of the participants' responses, including frequency and percentage, as well as the results of Chi-square tests are presented in the following table.

Table 3: Results of the participants' survey

	Responses			Total
	High	Average	Low	
Teachers	20 (14 %)	20 (14 %)	95 (72 %)	135
Master students	100 (13 %)	150 (20 %)	470 (67%)	720
Doctoral students	60 (13 %)	90 (21%)	300 (67%)	450

As the results in the above table indicate (X^2=38.149, df=4, Sig.= .000), there is a significant difference between the participants' responses. 14% of the teachers stated that the ESP tests had high authenticity, 14% thought that the authenticity of these tests was average, and 72% argued that the authenticity of such tests was low. Almost 33% of the master students argued that authenticity of these tests was average or high, whereas, 67% stated that the ESP tests had low authenticity. Moreover, 31% percent of doctoral students believed that ESP tests had high or average authenticity; on the contrary, 67% argued that these tests had low authenticity.

4.2 Qualitative method and outcomes

In order to determine the perceived authenticity of the ESP tests, we interviewed 10 subject specialists, 15 PhD students and 30 master students of the aforementioned disciplines. The participants were asked to comment: whether authenticity is needed or not; whether these tests are needed to select the most qualified candidates or not, and whether it is necessary to make a change in the use of ESP tests or not. The transcribed semi-structured interviews were content analyzed. Interesting commonalities among the interviewees were: 'ESP tests are needed but they should not serve as selection tests. Almost all stated that the students' scores on these tests should serve as indicators of the language areas in which they may lack

competence and therefore need further instruction'. That is, the test-takers' performance on these tests should not influence their entry into post-graduate studies. The next extracted theme was 'that authenticity is nice but absent in these tests'. The participants emphasized that, regardless of the purpose, 'ESP tests should cover all language skills needed by the learners'. Another finding was that ESP tests, because of the testing method (multiple-choice items) 'do not measure the test-takers' real abilities in the use of academic language in target situations'.

5. Discussion

In Iran, in the last three decades there has been a steady growth in the number of language courses for ESP/EAP, which aim at offering students necessary skills and knowledge to be able to function efficiently in a specific academic or professional context. Over the past decades, language tutors have been developing steadily not only course materials, but also specific-purpose tests as part of entrance examinations to master and doctoral studies. A distinguishing characteristic of ESP tests is authenticity. In this study, we tried to identify the perceived authenticity of ESP tests through different approaches and techniques. An important common outcome of these approaches is that the ESP tests administered in 2006 and 2007 lacked this characteristic.

In terms of task authenticity, the results of the study in Table 1 as well as the content analysis of the sample tests in Appendix I indicate that target language use situation tasks and ESP tests tasks do not match. As the characteristics of target language use situations, rubrics, input, expected response, and the interaction between input and response, are not similar to those of ESP test tasks, it could be strongly argued that the ESP tests administered to select MA/MSc and PhD candidates lack the required degree of authenticity.

In addition to task authenticity, correspondence between learners' language needs and test tasks is another distinguishing characteristic of ESP/EAP tests. To determine this correspondence, learners' language needs were compared with the types of tasks and the skills measured by these ESP tests. The results indicated that no significant correspondence was identified. Therefore, it could be said that there is a great difference between what language learners need and what is measured by ESP tests. Moreover, about 75% of the respondents indicated that ESP tests are not representative of the learners' language needs. They argued that none of the tasks of these tests correspond to TLU situation tasks, nor do they measure the learners' abilities to use language for their own academic needs.

Another influential factor in authenticity of language tests is test purpose which is assumed to have significant impacts on authenticity of ESP tests (Qichun, 2005). The single most important consideration in the development of language tests is the purpose which particular tests are intended to serve. The two major uses of tests are: to make decisions about the achievements of the test-takers and select the most qualified candidates to enter universities; as indicators of the learners' abilities to use language in TLU domains and the areas in which learners may lack competency significantly and need remediation. The findings of this study indicated that there is consensus among the two different groups of the interviewed participants (master and doctoral students) that ESP tests administered to the learners of the aforementioned disciplines should not serve as a criterion to evaluate the test-takers' achievements, or to select the qualified ones and reject the others. Instead, these tests should serve as an indicator of the areas in which master and doctoral students may be weak and need to learn. The participants also indicated their scores on these tests should serve as an indicator of their learning problems. Therefore, in line with the findings, it could be said that a change in the purpose of the ESP test is really necessary. It is also essential that the students' scores should not influence their entry or non-entry into universities. However, the students'

scores on their content subjects should serve as an indicator of whether the candidates are qualified or not, and ESP tests should be administered after the candidates are screened as an indicator of their capability in ESP/EAP or to see whether they need further instruction.

Based on the review of the literature, (Bachman &Palmer, 1996; Qichun, 2005; Trantik, 2008), it could be argued that authenticity is a critical test quality because it relates the test tasks to the domain to which the test administrators want their score interpretations to generalize. Authenticity, therefore, provides a means for investigating the extent to which score interpretations could be generalized beyond ESP learners' performance on the test to language use in TLU domain, or to other non-test real academic situations. Such definition of authenticity links it to construct validity (Alibakkshi, Kiani, & Akbari, 2010; Backman & Palmer, 1996).

On the other hand, in conformity with the view of Wu and Stansfield (2001), it is deemed clear that if ESP tests are to be considered valid, the authenticity of their language and tasks must be verified and the test task must be the closest equivalent of what one encounters in specific work. On the contrary, ESP tests administered at Iranian universities, due to lack of authenticity, would certainly lead to their construct underrepresentation. It could also be said that these tests cannot have positive washback impacts on the learning and teaching process of ESP in Iranian universities, nor can test users generalize their score interpretations beyond ESP learners' performance on the tests to target academic situations.

6. Conclusion

The primary goal of ESP tests is to obtain information about the learner's specific purpose language ability. This information is often very useful and at times even essential. Globalization and rapid international development of science and technology encourage mobility of students across the English-speaking countries and beyond. In order to be able to

take up study or work opportunities, knowledge of foreign language is essential. At present, it is increasingly important not only to be able to use a foreign language, but also to be able to demonstrate that one can use it at the level required by employers, schools, or universities. Considering this, ESP examinations are gaining more and more attention, and are becoming an indispensable tool in the modern educational system.

The overall ESP test quality or usefulness is determined by the appropriate balance among all qualities of tests, generalizability, washback, topical knowledge, and authenticity. Although each quality is directed and controlled by a theory behind it, only the authenticity factor can manifest itself tangibly in test papers. Moreover, authentic assessment aims to evaluate students' abilities in 'real-world' contexts. In other words, students learn how to apply their skills to authentic tasks and projects. Authentic assessment does not encourage rote-learning or passive test-taking. Instead, it focuses on students' analytical skills, ability to integrate what they learn, creativity, ability to work collaboratively, and written and oral expression skills. It values the learning process as much as the finished product**.** Therefore, if test constructors are able to keep a firm grip of authenticity, they can guarantee the intangible overall test quality. As we know, authenticity involves the correspondence between characteristics of test tasks and the features of TLU tasks. It is highly recommended that test constructors take a systematic analysis and consideration of authenticity starting from the designing stage where test tasks are selected and constructed until the operational stage, so that they do not miss the best time to control authenticity. To do so, not only do they need to carefully analyze learners' specific/academic needs, but they also need to analyze features of specific target language use situation tasks. The authenticity of tests also depends to a great extent on test constructors' understanding and knowledge of the test-takers' characteristics, therefore the tests which go beyond the scope of test-takers' topical knowledge and abilities

fail to be authentic. Lack of authenticity would certainly lead to less reliable tests, and the inferences drawn to evaluate and make decisions about the test-takers are not reliable.

References

Alderson, J.C. (1981). Reaction to Morrow paper. In J.C. Alderson & A. Hughes (Ed.), *Issues in language testing* (ELT Documents No.111, pp. 45-54). London: The British Council.

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

Bachman, L. (1991). What does language testing have to offer? *TESOL Quarterly, 25*, 671–704.

Bachman, L and Palmer, A. (1996*). Language testing in practice*. Oxford: Oxford University Press.

Breen, M. (1985). Authenticity in the language classroom. *Applied Linguistics, 6*, 60–70.

Carroll, B.J. (1980). *Testing communicative performance*. London: Pergamon.

Carlson, D. 1991. Changing the face of testing in California. *California Curriculum News Report*, 16.3.

Davies, A. (1984). Validating three tests of English language proficiency. *Language Testing, 1*, 50–69.

Davies, A. (1988). Communicative language testing. In A. Hughes (Ed.), *Testing English for university study* (ELT Documents No.127). Oxford: Modern English Publications.

Douglas, D. (2000).*Assessing language for specific purposes*. Cambridge: Cambridge University Press.

Doye, P. (1991). Authenticity in foreign language testing. In S. Anivan (Ed.), *Current developments in language testing* (pp. 103–10). Singapore: SEAMEO Regional Language Centre.

Farhadi, H. (2008). *Needs analysis project*. Tehran: SAMT publications.

Fulcher, G., (2000). The `communicative' legacy in language testing. *System, 28*, 483- 497

Alibakkshi, G., Kiani, G. R., & Akbari, R.(2010). Assessment in ESP/EAP: development and construct validity. Unpublished PhD dissertation, Tarbiat Modares University, Tehran, Iran.

Lewkowicz, J. (2000). Authenticity in language testing: some outstanding questions. *Language testing*, *17*(1), 43–64.

Lynch, A.J. (1982). 'Authenticity' in language teaching: Some implications for the design of listening materials. *British Journal of Language Teaching, 20*, 9–16.

Morrow, K. (1983). The Royal Society of Arts Examinations in the Communicative Use of English as a foreign Language. In R. Jordan (Ed.), *Case studies in ELT*. (pp. 102-107) London: Collins ELT.

Morrow, K. (1991). Evaluating communicative tests. In S. Anivan (Ed.), *Current developments in language testing* (pp. 111–118). Singapore: SEAMEO Regional Language Centre.

Qichun, Z. (2005). The analysis of authenticity in language tests. *CELEA Journal*, *28*(1), 122-125.

Spolsky, B. (1985). The limits of authenticity in language testing. *Language Testing,* 2, 31–40.

Trantik, A.(2008). Key issues in testing English for specific purposes. *Scripta Manent*,4(1) 3-13

Widdowson, H. (1978). *Teaching language as communication*. Oxford: Oxford University Press.

Widdowson, H. (1990). *Aspects of language teaching*. Oxford: Oxford University Press.

Widdowson, H. (1994) The ownership of English. *TESOL Quarterly, 28*, 377–89.

Widdowson, H. (1998) Context, community and authentic language. Paper presented at TESOL Annual Convention, Seattle, 17–21 March 1998.

Wu, W. M., &. Stansfield, C, W. (2001). Towards authenticity of task in test development. *Language Testing*, *18*(2), 187-206.

Appendix I

Specific English test part of master entrance examination: geography discipline 2007

Directions: read the following three passages and choose the best item a, b, c, or d. then mark the correct answer on your answer sheet.

PASSAGE 1:

Much of the world is increasingly urbanized. Before 1850 no country could claim to be predominantly urbanized and at the turn of the century, only Great Britain could be regarded as such. Urbanization is the process whereby society is transformed from an essentially rural one to a predominantly urban one. It's most visible expression in the landscape is the growth of cities and an increase in their number, size and importance. Also, urbanization process can be closely allied with economic colonization. Urbanization however is not just the growth of cities but rather it is a complex change in economic, social and political thinking. As cities grow, new definitions have been introduced for census purposes in which the concept of the "extended city" is commonly used. Definitions such as the urban area based on the physical extent of the built-up area and minimum requirements of population size and density as well as the metropolitan area, based on urban population size and commute patterns as a measure of spatial integration give us a much better indication of the population size of cities than the use of political entities like the municipality. One of the definitions that have raised much interest is that based on the concept of urban field which is a new form of urban habitat of relatively low density involving a good transportation system and a broad array of economic, social and recreational opportunities.

1- According to the passage, most of the world ----------.

a) Had changed into cities before I 850

b) Country sides have changed into cities and towns

c) Cities and towns had grown bigger before 1850

d) Population has suffered economic colonization

2- What does 'whereby' in line 4 mean?

a) by which b) in which c) in the place where d) by the place where

2- The growth of cities in number, size, and importance-------economic colonization.

a) must be related to b) should be connected with

c) is sometimes connected with d) is always related to

3- On the basis of the passage, urbanization is ----------.

a) just the most visible expression in the landscape

b) the mere increase of cities in number, size, and importance

c) a complex change in economic, social and political colonization

d) a complicated development in economic, social and political thinking

4- The author concludes that the use of political entities like the municipality ----------.

a) involves a good transportation system

b) is less indicative of the population size of cities

c) is more indicative of the population size of cities

d) involves a bread array of economic, social and recreational opportunities

PASSAGE 2:

Although health care is a public good, it is not 'pure'. In other words, it is not equally available

to all individuals. The geography of health care provision has much to do with this. Demand for public services is continuously distributed across space, broadly in accordance with the distribution of population, but these services are only provided at discrete locations. Inevitably therefore, there will be inequalities of access in terms of the practicality of using services, transport costs, travel times and so on. Geographical or 'locational' factors (e.g. physical proximity, travel time) are not the only aspects which influence access to health care. Other types (or dimensions) of accessibility to health care except for geographical (or spatial) are social, financial and functional. *Social* accessibility to health care depends on race (like separate hospitals for white and black people), age, sex and other social characteristics of individuals, important here is also relationship between patient and the doctor. *Financial* depends upon the price of a particular health care and *functional* reflects the amount and structure of provided services. This can vary among different countries or regions of the world.

5- What does the passage mainly discuss?

a) Health Care b) Accessibility to Health Care

c) Social Accessibility to Health Care d) Location Factors in Health Care

6- Demand for public services is continuously distributed across space ----------.

a) only according to the distribution of population

b) mostly on the basis of the distribution of population

c) only based upon the geography of health care provision

d) mostly based upon the geography of health care provision

7- Such factors as ---------- are included among the geographical factors which ---------.

a) physical proximity and travel time - do affect accessibility to health care

b) physical proximity and travel time - do not affect accessibility to health care

c) using services, transport costs and travel times - are social, financial and functional

d) using services, transport costs and travel times - are not social financial and functional

8- Based upon the passage, there are ---------- types of accessibility to health care, namely social, financial ----------.

a) three - and physical b) three - and functional

c) four -, spatial, and physical d) four -, functional, and spatial

9- What docs 'this' in line 14 refer to?

a) amount and structure of services b) the price of a particular health care

c) functional accessibility to health care d) the relationship between patient and the doctor

PASSAGE 3:

The term political geography was first used by Friedrich Ratzel in his book 'Politische Geographie', published in German in 1897. Geopolitics was then coined by the Swede Rudolf Kejel1.

The discipline gained attention largely through the work of Sir Halford Mackinder in England and his formulation of the Heartland Theory in 1904. This theory involved concepts diametrically opposed to the notion of Alfred Thayer Mahan about the significance of navies (Mahan coined the term *sea power)* in world conflict. The Heartland theory, on the other hand, hypothesized the possibility for a huge empire to be brought into existence which didn't need to use coastal or transoceanic transport to supply its military industrial complex, and that this empire could not be defeated by all the rest of the world coalitional against it.

The Heartland Theory depicted a world divided into a *Heartland* (Eastern Europe/Western

Russia); *World Island* (Eurasia and Africa); *Peripheral Islands* (British Isles, Japan, Indonesia and Australia) and *New World* (The Americas). Mackinder claimed that whoever controlled the Heartland would have control' of the world. He used this warning to politically influence events such as the Treaty of Versailles, where buffer states were created between the USSR and Germany, to prevent either of the them controlling the Heartland.

10- Which statement is true about 'geopolitics'?

a) Rudolf Kejel used 'geopolitics' for the first time.

b) The Swede Rudolf Kejel made coins from metal using 'geopolitics'.

c) Friedrich Ratzel used the word 'geopolitics', which no one had already used.

d) The German Friedrich Ratzel first used 'geopolitics' in his book 'Politische Geographie'.

11- the heartland theory formulated by Halford Mackinder ---------- Alfred Thayer Mahan's notion.

a) rarely opposed to b) in some ways opposed to

c) was completely different from d) was partially different from

12- Halford Mackinder believed that the rest of the world ----------.

a) needed to use coastai or transoceanic transport to defeat it.

b) didn't need to use coastal or transoceanic transport to overcome it

c) could defeat the huge empire if they established a coalition against it

d) could overcome the huge empire even if they established a coalition against it

13- What does the word 'depicted' in line 12 mean?

a) portrayed b) established c) conquered d) created

14- Mackinder brought about the Heartland Theory in order to ----------.

a) affect events such as the Treaty of Versailles

b) create events such as the Treaty of Versailles

c) weaken buffer states between the USSR and Germany

d) establish buffer states between the USSR and Germany

Specific English test part of master entrance examination: Persian literature discipline 2007

Directions: read the following three passages and choose the best item a, b, c, or d. then mark the correct answer on your answer sheet.

With life as short as it is, with so many pressing demands on our time, with books of information, instruction, and discussion waiting to be read, why should we spend precious time on works of imagination? The eternal answers to this question are two: enjoyment and understanding.

Since the invention of language, men have taken pleasure in following and participating in the imaginary adventures and imaginary experiences of imaginary people. Whatever without causing harm - serves to make life less tedious, to make the hours pass more quickly and pleasurably, surely needs nothing else to recommend it. Enjoyment - and ever more enjoyment - is the first aim and justification of reading fiction. But, unless fiction gives something more than pleasure, it hardly justifies itself as a subject of college study. Unless it expands or refines our minds or quickens our sense of life, its value is not greater than that of ping-pong or any other games. To have a compelling claim on our attention, it must yield not only enjoyment but understanding.

The experience of men through the ages is that literature may furnish such understanding and do so effectively that the depiction of imagined experiences can provide authentic insights. Fiction may be classified into two broad categories: literature of Escape and literature of Interpretation. Escape literature is written purely for entertainment - to help us pass the time agreeably. Interpretive literature is written to broaden our awareness of life. Escape literature takes us away from the real world: it enables us temporarily to forget our troubles.

Interpretive literature takes us, through the imagination, deeper into the real word: it enables us to understand our troubles. Escape literature has as its only object pleasure. Interpretive literature has as its object pleasure plus understanding.

1- What does the passage mainly discuss?

1) Authentic insights 2) Depiction of imagined experiences

3) The reason for reading fiction 4) The reason for reading literature

2- According to the author, men have enjoyed imaginary stories since the _____.

1) beginning of human life 2) beginning of the universe

3) development of literature 4) development of language

3- What does "tedious" in line 7 refer to?

1) boring 2) lively 3) interesting 4) exciting

4- The first goal of reading fiction is <u>not</u> _____ .

1) amusement 2) displeasure 3) entertainment 4) recreation

5- Interpretive literature is <u>not</u> developed to _____ .

1) deepen our insight . 2) limit our knowledge

3) increase our appreciation 4) sharpen our understanding

Literature is writing in prose or verse that has excellence of form or expression. Usually, literature presents ideas of permanent or universal interest. The word is often used to describe groups of writings - such as 'Children's Literature' or those of a special place or time, such as 'Russian Literature' and so on. Generally, the term means the widely recognized and long-lasting written material that can be called ART.

One chief form of literature is the Novel, a long work that often tells a complex story about one or more people. In telling it, the author may create exciting settings, realistic characters,

and elaborate situations in which a great deal happens. The short story is much less complicated, its characters are fewer and not as completely developed, and their actions are less wide-ranging.

The Drama is another major form of literature, one which tells its story in speech and action. For this reason, plays can be performed before audiences and can be enjoyed by many more people at once than the novel or short story.

Poetry is an important branch of literature. It can tell a story, or it can create moods and reveal feelings. In some cases, poetic forms are applied to other areas - many of William Shakespeare's plays are written in poetic form.

The second major division of literature is nonfiction, which tells about real people, places, events, or ideas. Nonfiction includes histories, or reports of events in the past. Essays are often personal statements or arguments on particular subjects, and expositions are discussions or descriptions of ideas or places. A nonfiction writer sometimes moves toward fiction in biography, a story of somebody's life. In telling of a person's thoughts and actions, some biographers tend to use imagination somewhat more than in other writing.

Much of what is written everyday is prose nonfiction. Newspapers are written in this form, as are magazines and student's textbooks. But these do not fit into the class of literature because they generally lack permanent or universal value.

6- Literature is writing in prose or verse that presents _____ ideas.

1) widely limited 2) widely acceptable

3) temporarily limited 4) temporarily acceptable

7- A short story is more _____ and the characters in a novel is more _____ developed.

1) elaborate, partially 2) sophisticated, partially

3) straightforward, fully 4) complicated, fully

8- According to the passage, play is a form of _____ like a(n) _____.

1) non-fiction, novel 2) non-fiction, essay

3) fiction, essay or history 4) fiction, novel or short story

9- Poetic forms can be used to create _____.

1) novels and short stories 2) plays and novels

3) histories or short reports 4) novels, plays and short stories

10- In non-fiction literature, the writer _____ towards fiction

1) never moves 2) always inclines

3) often inclines 4) occasionally moves

Analysis attempts to find truth. The process of analysis is to divide a problem into various parts, for although a whole object is difficult to comprehend at one glance, the parts may be examined more easily, and their natures, functions, and interrelationships may be more fully understood when they are examined one by one. For example, if you have the problem in chemical qualitative analysis of discovering the elements in a chemical solution, you can make only one test on the solution at a time, because if you tried to make all your tests at once you would not be able to control or distinguish your results.

The analysis of literature is based on the same truth. Although the work you have read is an entirety, you must ask separate questions in order to discover its full meaning and to appreciate it fully. You could not talk about everything in Paradise Lost at once, for example, without being guilty of the greatest superficiality. It is better, in your discussion, to narrow the scope of your topic by talking about the diction, epic conventions, theology, or dramatic action. An attempt to discuss everything at once would most probably distort some things and omit others; results of this sort of investigation are usually wrong or misleading. Truth, however, can emerge only if all possibilities are considered. So your problem in making an

analysis is to make the subject small enough so that you can go deeply into it. In other words, you can write a good theme about an almost limited part or aspect of the work you have read; but you would find it impossible to discuss everything unless your analysis grew to the length of a book.

A serious objection sometimes arises about literary analysis. Although scientific analysis is necessary, it is said that too much literary analysis "spoils" appreciation of a work, or that in making an analysis, you "murder" literature as you analyze it.

11- What is the main goal of the process of analysis?

1) to discover facts 2) to discover interrelationships

3) to examine something easily 4) to examine something fully

12- Unless you make only one test at a time, you will _____ the results.

1) be able to describe 2) be able to control

3) not be able to control 4) not be able to describe

13- if you talk about everything in Paradise Lost at once, you will _____.

1) be blamed 2) be criminal

3) not be blamed 4) not be criminal

14- The results of discussing everything at once are <u>not</u> usually _____.

1) distorting 2) illuminating

3) misleading 4) superficial

15- According to the author, if your analysis didn't become as big as a book, it would be impossible for you _____.

1) to monitor everything 2) to control the results

3) to distinguish the results 4) to discuss everything

The Asian ESP Journal

The impact of personality traits on the writing performance of Iranian EFL learners

Mohammad Mehdi Soleimani
Islamic Azad University, Karaj Branch
Mahmoud Reza Daryabari
Allame Tabatabaee University

Biodata

Mohammad Mehdi Soleimani is a PhD candidate in Applied Linguistics at Islamic Azad University, Science and Research Campus. He is currently the Head of English Department at Islamic Azad University, Karaj Branch. His research interests include language teaching and testing, interlanguage pragmatics, form-focused instruction, individual learner differences, and language learning strategies.

Mahmoud Reza Daryabari holds an MA in TEFL from Allame Tabatabaee University, Tehran, Iran. He has extensive experience of teaching English as a foreign language. His current fields of interest include language teaching and testing.

Abstract

This study investigates the impact of language learners' personality on the successful execution of academically-oriented language tasks. The rationale for undertaking this inquiry was to test whether the personality type, i.e., introversion vs. extroversion, had an impact on

the learners' success at accomplishing writing tasks. Since writing skill depends highly upon the learners' cognitive academic language proficiency (CALP), the students' composition scores were identified as an indication of their ability to execute academically-oriented language tasks. To achieve the objectives of the study, a total of 92 Iranian EFL students took a TOEFL test of English language proficiency. Based on TOEFL scores, 63 students were selected to take the Eysenck Personality Questionnaire (EPQ) and to write a composition in English. The results revealed that extroverted language learners outperformed introverts on doing academically-oriented language tasks. This finding contradicted Griffiths's (1991) claim that introverted learners will do better than extroverts at developing cognitive academic language abilities.

Keywords: Introversion; extroversion; gender; CALP; BICS; writing ability

1. Introduction

The ways in which language learners differ are potentially varied. These learner differences may stem from various sources that would reflect a whole range of variables relating to the cognitive, affective, and social aspects of human beings (Ellis, 1994). Language teachers, however, seem to ignore this well-attested truth that each individual learner is unique. This ignorance surfaces when language teachers fill identical prescriptions for language learners without paying due attention to these inherent individual learner differences. Addressing this concern, Kinsella (1995) cautions language teachers against adopting such unyielding attitudes towards language learners and the way they approach the learning process. He argues that:

> Without a fundamental awareness of individual preferences, it is easy to believe that the way we learn is the most efficient way;

> consequently we will bias our teaching in favor of those who approach learning in much the same way as we do (p.180).

It is thus highly recommended that language teachers consider these suggestions with greater appreciation in order to emancipate their teaching style from the shackles of such biased practices.

2. Introversion and extroversion: A short review of previous studies

Second language acquisition (SLA) research has revealed that success or failure in learning languages is due not only to cognitively correlated issues but also to affective, motivational, demographic, and personality factors (Carroll, 1990; Ehrman & Oxford, 1990, 1995; Oxford, 1992; Oxford & Ehrman, 1993; Skehan, 1989). Out of these factors, personality traits have drawn a substantial deal of attention in SLA research and studies based on these have made extensive contributions to our understanding of language learning processes.

Currently, the research domain that focuses on language learners' personality traits is dominated by Eysenck's three component construct and the Big Five model (Dornyei, 2005). As Dornyei argues, Eysenck's model aims at contrasting: 1) extroversion with introversion, 2) neuroticism and emotionality with emotional stability, and 3) psychoticism and tough-mindedness with tender-mindedness. The Big Five construct, on the other hand, retains Eysenck's initial dimension, however it replaces psychoticism with three additional dimensions of conscientiousness, agreeableness, and openness to experience. Dornyei (2006) considers this overlap between the two constructs as an advantage for non-psychologist researchers because it provides them with an easier means for investigating language learners' latent traits.

Thus far, the most researched personality traits in language studies have been listed as the extroversion/introversion dichotomy (Dornyei, 2005; Ellis, 2004). Brown (2000) defines the terms as follows:

> Extroversion is the extent to which a person has a deep-seated need to receive ego enhancement, self-esteem, and a sense of wholeness from other people as opposed to receiving that affirmation from within oneself. Introversion, on the other hand, is the extent to which a person derives a sense of wholeness and fulfillment apart from a reflection of this self from other people (p. 155).

Dewaele and Furnham (1999) relate the apparent discrepancies between extroverts and introverts to biological arousal levels in the two groups. The authors argue that "extroverts are under-aroused and introverts are over-aroused" (p.512). Considering the fact that individuals operate ideally within a moderate arousal level, Dewaele and Furnham argue that extroverts look for external stimulation to reach optimal arousal levels, while introverts try to avoid such stimulations. In other words, introverts would evade such states because they are prone to suffer from arousal levels that exceed their optimal tolerance.

There are two major hypotheses regarding the relationship between the extroversion/introversion dichotomy and the process of second language acquisition (Ellis, 1994; 2008). The first hypothesis suggests that extroverts perform better when they encounter learning tasks that engage their basic interpersonal communicative skills (BICS). The rationale for this hypothesis, as Ellis (2008) states, is that the learners' sociability, which is an essential feature of extroversion, provides more opportunities for practice, more input, and in turn results in greater success in communicating in an L2. The second hypothesis, on the other hand, indicates that introverts outperform extroverts in accomplishing tasks that primarily rely on language learners' cognitive and academic language proficiency (CALP). The rationale for

such a hypothesis is that introverted learners prefer to immerse themselves in reading and writing than integrating with the surroundings, therefore developing their cognitive and academic abilities (Ellis, 2008).

The evidence linking extroversion to the acquisition of BICS is fairly substantial. Following a thorough examination of twelve studies that addressed the issue, Strong (1983) found a meaningful relationship between extroversion and the acquisition of BICS in at least six of the reviewed studies. The second hypothesis, however, has received meager support from the literature. Strong (1983) reported an insignificant relationship between introversion and CALP. Busch (1982) also failed to find any significant relationship between the two variables. These findings were further supported by Carrel, Prince and Astika (1996) who found an insignificant relationship between introversion and the learners' writing score.

Skehan (1989) disregards this paucity of evidence in support of the second hypothesis and attaches a great degree of significance to extroversion by arguing that extroverted learners are just as successful as introverts in accomplishing cognitive and academic language tasks. To assess the validity of Skehan's argument, the present researchers tried to examine any possible sources of interrelationship between the introversion/extroversion dimension of learner personality and success at accomplishing writing tasks, where successful performance is believed to depend on the learners' cognitive and academic language abilities. To this end, the following research questions were formulated:

1. Does learner personality affect performance on writing tasks?
2. Does gender play a significant role in the writing performance of language learners?
3. Does the interaction effect between the learners' gender and personality influence their writing performance?

3. Method

3.1 Participants

A total of ninety two undergraduate EFL students participated in the initial phase of the study. These participants were randomly selected from among the sophomore and junior English majors of Karaj Azad University in Iran. The participants were further required to take a TOFEL test to provide the researchers with a homogeneous sample of language learners. Finally, on the basis of TOEFL scores, sixty three students were selected to take part in later stages of this study.

3.2 Research Instruments

The following research instruments were used in this study:

1. In order to measure the participants' language proficiency, a standard version of TOEFL test was administered. The reliability of the instrument for the sample, using KR-21, was found to be 0.80.
2. In order to identify the learners' personality traits, Eysenck Personality Questionnaire (EPQ) was used. As discussed earlier, Eysenck's three component construct is widely accepted by scholars and the questionnaire is now widely used by researchers. The modified version of EPQ included twenty four yes-no questions to be completed by the participants. The calculated reliability of the instrument for the sample was found to be 0.85.
3. In order to measure the participants' performance on academically-oriented language tasks, a writing test was administered. For this purpose, the subjects were required to write a composition of about 300 to 400 words.

3.3 Procedure

As discussed above, a total of ninety two Iranian university students took a TOEFL test. The scoring procedure was based on TOEFL standards. To select a sample of homogeneous students, the researchers set a cut-off score of 450 as a criterion for selecting sample participants. The sixty three students who could meet the criterion score were further required to take the Eysenck Personality Questionnaire. The participants were also asked to write a composition of about 300 to 400 words on the following topic: "An English-speaking friend is coming to stay with you for the weekend. Write a letter telling him/her about three ideas you have for his/her visit." The students had ninety minutes to plan, pre-write, draft, revise, and edit their compositions before returning them to the researchers. Despite existing reservations about the timed in-class writing samples, many scholars still regard the technique as one of the most reliable and most valid assessment tools available to researchers (Foster, 1983).

To increase the scoring reliability, the researchers first devised a criterion or a marking scheme which is generally defined as a "document indicating the explicit criteria against which candidates' performance will be judged" (Murphy, 1979, p.80). This marking scheme was used with the purpose of enabling the examiner(s) to relate particular marks to answers of specified qualities. In this regard, a marking scheme based on Brown and Bailey (1984) was adapted. What follows is a representation of various components of the marking scheme that was used in this study: 1) Organization, 2) Logical development of ideas, 3) Structure, 4) Punctuation and mechanics, and 5) Style and quality of expression. A total of four points was allocated to each component of the marking scheme, and the participants' overall composition score was considered to be the sum of scores assigned to these five components. The researchers also had the compositions scored by three different raters. Once the raters scored the compositions, the inter-rater reliability index was calculated using coefficient alpha formula which showed a high index of 0.70.

4. Design and data analysis

In this study, the extroversion/introversion dichotomy of personality traits was treated as the independent variable along with the participants' writing performance that served as the dependent variable. Gender and proficiency were also included in the study as the moderator and control variables respectively. To answer the first two research questions, the researchers conducted independent samples t-test. However, due to the nature of the third research question the researchers had to utilize a two-way analysis of variance (ANOVA) to identify the nature and the direction of interaction between the independent variables (i.e., personality traits and gender) and the dependent variable. The rationale for running a two-way analysis of variance was to explore the effects of following independent variables on the writing performance of the sample participants:

1. The effect of personality traits (Factor A): introversion vs. extroversion
2. The effect of gender factor (Factor B): male vs. female
3. The interaction effect of personality and gender (Factor A x B)

5. Results and Discussion

The following table provides some preliminary descriptive statistics on the writing performance of introverted as opposed to extroverted language learners who participated in this study.

Table 1. Group Statistics

	Personality	N	Mean	Std. Deviation	Std. Error Mean
Writing	**Extrovert**	32	16.5228	.84255	.14894
	Introvert	31	16.0600	.83672	.15028

Group statistics indicate that introverts display a somewhat lower mean score for writing than extroverts (i.e., 16 versus 17). No conclusions, however, can be drawn from the displayed data in table 1. The key question that remains unanswered is whether this existing difference between the two groups' performance can reach a significant statistical level to reject the first null hypothesis. To answer this key question, the researchers conducted independent samples t-test to compare the composition scores of the two groups. The t-test results warranted the rejection of the first null hypothesis and in turn confirmed the researchers' alternative hypothesis. This finding can be explained by referring to the displayed data in figure 2 for either the t-value or the probability level. We can reject the null hypothesis because our t-value is greater than 1.98 and our obtained significance level, i.e., 0.03, is smaller than the conventional 0.05 magnitude. The researchers also calculated the effect-size using Cohen's d formula and obtained a moderate effect-size of 0.51 for the sample.

Figure 2. Independent sample t-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**
Equal variances assumed	.010	.923	2.187	61	.033*	.46281	.21161	.03967	.88595
Equal variances not assumed			2.187	60.96	.033	.46281	.21159	.03972	.88591

To address the second research question, the researchers also conducted independent samples t-test to identify the impact of moderator variable on the writing performance of the learners. Group statistics, as displayed in table 3, showed a slight mean score variation in favor of female participants.

Table 3. Group statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Writing	**Male**	29	15.9559	.88936	.16515
	Female	34	16.5844	.73962	.12684

Further t-test analysis of the data provided evidentiary support for the rejection of the second null hypothesis and thus verifying the researchers' alternative hypothesis. Levene's test showed no significant mean difference between the male and female groups (i.e., p> 0.05), thus sustaining the equal variance assumption. The second null hypothesis was rejected based on the fact that the probability figure – marked as 'Sig. (2 tailed)' in table 4 – showed a smaller figure than the conventional 0.05 value. The magnitude of the difference between the means also showed a moderate effect-size of 0.75 using Cohen's d formula.

Table 4. Independent samples t-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
Writing	**Equal variance**	.729	.397	-3.063	61	.003*	-.62855	.20520	-1.03887	-.21823
	Unequal variance			-3.018	54.644	.004	-.62855	.20824	-1.04593	-.21116

To investigate the third research question, the researchers conducted a two-way analysis of variance to determine the interaction effect of gender and personality on the writing performance of the research participants. The SPSS output for the ANOVA test is presented in table 5. The data shows no significant interaction effect of personality and gender on the

dependent variable because the significance level exceeds the 0.05 value (p> 0.05). Considering this finding, the third null hypothesis cannot be rejected. This finding indicates that the personality and gender factors are strong enough to override the interaction effect between the two variables.

Table 5. Two-way analysis of variance

Dependent Variable: Writing

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	9.277[a]	3	3.092	4.917	.004	.200
Intercept	16533.888	1	16533.888	2.629E4	.000	.998
Personality	3.046	1	3.046	4.843	.032	.076
Gender	5.792	1	5.792	9.210	.004	.135
Personality * Gender	.126	1	.126	.200	.656	.003
Error	37.106	59	.629			
Total	16774.748	63				
Corrected Total	46.383	62				

a. R Squared = .200 (Adjusted R Squared = .159)

6. Conclusion

The study provided some support for Skehan's (1989) assumption by confirming that extroverts outperform introverts in language skills whose mastery depends more profoundly on language learners' cognitive and academic abilities than their communicative skills. This finding contradicts the commonly held belief that introverts have an advantage in performing cognitive academic tasks (Griffiths, 1991). The study also confirmed that females outperform males in carrying out cognitive and academic tasks (Nyikos, 1990). The reason for the abundance of contradictory results that surround the research on personality traits might lie, in part, in 'methodological issues'. As Dewaele and Furnham (1999) argue:

> The majority of these studies have been performed by researchers working in different disciplines with different methodologies and

> hypotheses. It is not surprising, therefore, that misunderstandings and generalizations might arise from a perfunctory overview of this literature (Dewaele and Furnham, 1999, p. 518).

In conclusion, it should be borne in mind that our knowledge about the true nature of these variables is neither adequate nor comprehensive enough to authorize establishing clear-cut administrative decisions. Nonetheless, teachers' awareness of the existence of such individual differences can sensitize them to the varied reactions that learners might display as they face innovative teaching methods and learning materials, and would in turn help them adjust their teaching style based on their students' learning needs.

References

Brown, H.D. (2000). *Principles of language learning and teaching* (4th ed.). New York: Longman.

Brown, J. D., & Bailey, K. M. (1984). A categorical instrument for scoring second language writing skills, *Language Learning, 34*(4), 21-38.

Busch, D. (1982). Introversion-extraversion and the EFL proficiency of Japanese students. *Language Learning, 32*(1), 109-32.

Carrel, P. L., Prince, M. S., & Astica, C. C. (1996). Personality types and language in an EFL context. *Language Learning, 46*(1), 785-799.

Carroll, J. B. (1990). Cognitive abilities and foreign language aptitude: Then and now. In T. S. Parry, & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 11-29). NJ: Prentice Hall Regents.

Dewaele, J., & Furnham, A. (1999). Extraversion: The unsolved variable in applied linguistics research. *Language Learning*, *43*(3), 509-544.

Dornyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. NJ: Lawrence Elbaum Associates.

Dornyei, Z. (2006). Individual differences in second language acquisition. *AILA Review*, *19*, 42-68.

Ehrman, M. E., & Oxford, R. L. (1990). Adult language learning styles and strategies in an intensive training setting. *Modern Language Journal, 74*(3), 311,327

Ehrman, M. E., & Oxford, R. L. (1995). Cognition plus: Correlates of language learning success. *Modern Language Journal, 79*(1), 67-89.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R. (2004). Individual differences in second language learning. In A. Davis & C. Elder (Eds.). *The handbook of applied linguistics* (pp. 525-551). Oxford: Blackwell.

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

Foster, D. (1983). *A primer for writing teachers.* NJ: Boynton-cook.

Griffiths, R. (1991). Personality and second language learning: Theory, research, and practice. In E. Sadtano (Ed.). *Language acquisition and the second language classroom* (pp. 103-135). SEAMEO Regional Language Center.

Kinsella, K. (1995) Understanding and empowering diverse learners in ESL classrooms. In J. M. Reid (Ed.). *Learning styles in the ESUEFL classroom* (pp. 170-194). Boston: Heinle and Heinle.

Murphy, J. (1979). Removing the marks from examination scripts before re-marking them: Does it make any difference? *British Journal of Educational Psychology, 49*(1), 73–8.

Nyikos, M. (1990). Sex related differences in adult language learning. *The Modern Language Journal 74*(3), 273-278.

Oxford, R. L. (1992). Who are our students? A synthesis of foreign and second language research on individual differences. *TESL Canada Journal, 9,* 30-49.

Oxford, R. L., & Ehrman, M. E. (1993). Second language research on individual differences. *Annual Review of Applied Linguistics, 13*, 188-205.

Skehan, P. (1989). *Individual differences in second language learning.* London: Edward Arnold.

Strong, M. (1983). Social styles and the second language acquisition of Spanish-speaking kindergartners. *TESOL Quarterly, 17*(2), 241-258.

The Asian ESP Journal

Evaluation of SAMT ESP Textbooks for the Students of Medical Sciences

S.A. Razmjoo & Reza Raissi

Shiraz University

Biodata

Dr. Seyyed Ayatollah Razmjoo is an Assistant Professor in the Department of Foreign Languages and Linguistics at Shiraz University, Iran. His areas of interest are Testing, Research, Materials Development and Teaching Methodology. He has written articles and presented papers at both national and international conferences. His most famous book is Fundamental Foundations in TEFL.

Reza Raissi holds an MA in TEFL from Shiraz University, Iran. He is teaching general English courses in the Iranian Azad universities.

Abstract

English language teaching textbooks in general and ESP ones in particular are considered as one of the most important elements in any educational system. Because of their role and contribution to second language acquisition, ELT texts have become the major concern of many research studies, including the ones done in an EFL context such as Iran. This study aimed at describing the present state of SAMT (The Center for Studying and Compiling University Books in Humanities) ESP textbooks used in the Iranian universities of Medical Sciences from the viewpoints of students and instructors in order to provide a clear picture of

the current status of those textbooks. To fulfill the objectives, the following research questions were addressed in this study: (1) Are SAMT textbooks designed for the students of Medical Sciences appropriate in terms of the theoretical considerations, the organizational features and practical considerations, the content, the language skills, the vocabulary and the grammatical structure from the instructors' and students' viewpoints?; (2) What is the students' and instructors' overall impression of the textbooks?; and (3) Is there any significant difference among the students' and instructors' viewpoints? A quantitative design was used to evaluate ESP textbooks. Using availability sampling, 147 students and 30 instructors took part in the study. The questionnaire contained 55 specific criteria in six categories. The results indicate that instructors and students were not satisfied with most of the criteria which the SAMT ESP textbooks should fulfill. Moreover, the results show no significant difference between the students' and instructors' opinions.

Keywords: Textbook evaluation; ESP textbooks; SAMT Textbooks; ELT textbooks; EFL textbooks

1. Introduction

There are many vital components, such as teachers, learners, etc., in any educational system but the essential constituents of many ESL/EFL classroom and programs are the textbooks and instructional materials (contents) that are often used by language instructors. By the same token, Hutchinson and Torres (1994) suggest that "the textbook is an almost universal element of [English language] teaching. Millions of copies are sold every year, and numerous aid projects have been set up to produce them in [various] countries…. No teaching-learning situation, it seems, is complete until it has its relevant textbook" (p. 315).

Sheldon (1988) agrees with this observation and suggest that textbooks not only "represent the visible heart of any ELT program" (p. 237) but also offer considerable

advantages for both the student and the teacher when they are being used in the ESL/EFL classroom. Haycroft (1998), for example, suggests that one of the primary advantages of using textbooks is that they are psychologically essential for students since their progress and achievement can be measured concretely when we use them. Second, as Sheldon (1988) points, students often harbor expectations about using a textbook in their particular language classroom and program and believe that published materials have more credibility than teacher-generated or "in-house" materials. Third, as O'Neill (1982) and Ur (1996) indicate, textbooks are generally sensitive to students' needs, even if they are not designed specifically for them, they are efficient in terms of time and money and can and should allow for adaptation and improvisation. Fourth, textbooks yield a respectable return on investment and involve low lesson preparation time, whereas teacher-generated materials can be time, cost and quality defective. In this way, textbooks can reduce potential occupational overload and allow teachers the opportunity to spend their time undertaking more worthwhile pursuits (O'Neill, 1982; Sheldon, 1988). A fifth advantage, identified by Cunningsworth (1995), is the potential that textbooks have for serving several additional roles in the ELT curriculum. He argues that textbooks are an effective resource for self-directed learning and for presentation material, a source of ideas and activities, a reference source for students, a syllabus where they reflect pre-determined learning objectives, and support for less experienced teachers who have yet to gain in confidence. Although some theorists have alluded to the inherent danger of the inexperienced teacher who may use a textbook as a pedagogic crutch, such an over reliance may actually have the opposite effect of saving students from a teacher's deficiencies (O'Neill, 1982; Williams, 1983; Kitao & Kitao, 1997). Finally, Hutchinson and Torres (1994) point out that textbooks may play a pivotal role in innovation, suggesting that textbooks can support teachers through potentially disturbing and threatening change processes, demonstrate

new and/or untried methodologies, introduce change gradually, and create scaffolding upon which teachers can build a more creative methodology of their own.

While many of the aforementioned theorists and researchers are quick to point out the extensive benefits of using ESL/EFL textbooks, there are many other researchers and practitioners who do not necessarily accept this view and retain some well-founded reservations on the subject. Allwright (1981), for instance, has written a scathing commentary on the use of textbooks in the ELT classroom. He suggests that textbooks are too inflexible and generally reflect the pedagogic, psychological, and linguistic preferences and biases of their authors. Consequently, the educational methodology that a textbook promotes will influence the classroom setting by indirectly imposing external language objectives and learning constituents on students as well as potentially incongruent instructional paradigms on the teachers who use them. In this fashion, therefore, textbooks essentially determine and control the methods, processes and procedures of language teaching and learning. Moreover, the pedagogic principles that are often displayed in many textbooks may also be conflicting, contradictory, or even out-dated, depending on the capitalizing interests and exploitations of the sponsoring agent.

More recently, textbooks have been criticised for their inherent social and cultural biases. Researchers such as Porreca (1984), Florent and Walter (1989), Clarke and Clarke (1990), Carrell and Korwitz (1994), and Renner (1997) have demonstrated that many EFL/ESL textbooks still contain rampant examples of gender bias, sexism, and stereotyping. They describe such gender-related inequities as relative invisibility of female characters, the unrealistic and sexist portrayals of both men and women, stereotypes involving social roles, occupations, relationships and actions, as well as linguistic biases such as 'gendered' English and sexist language. Findings such as these have led researchers to believe that the continuing prevalence of sexism and gender stereotypes in many EFL/ESL textbooks may reflect the

unequal power relationships that still exist between the sexes in many cultures, the prolonged marginalization of females, and the misrepresentations of writers with social attitudes that are incongruent with the present-day realities of the target language culture (Sunderland, 1992; Renner, 1997).

Other theorists such as Prodromou (1988) and Alptekin (1993) have focused on the use of the target language culture as a vehicle for teaching the language in textbooks and suggest that it is not really possible to teach a language without embedding it in its cultural base. They argue that such a process inevitably forces learners to express themselves within a culture of which they have scarcely any experience and this may result in alienation, stereotyping, or even reluctance or resistance to learning.

Phillipson (1992) is also wary of the complex relationship between language textbooks and the target language culture but he sees the promotion of Western' (British) global textbooks as government-backed enterprises with both an economic as well as an ideological agenda. Gray (2000), on the other hand, has defended the socio-cultural components of many textbooks by saying that English language textbooks are actually ambassadorial cultural artifacts and students should not only critically engage their textbooks but also view them as more than mere linguistic objects. In this way, he argues, learners will improve their language skills by using their textbooks as useful instruments for provoking discussion, cultural debate, and a two-way flow of information. Clearly there is no consensus on this issue at this particular time and this would seem to warrant some degree of caution when using these types of books in certain teaching and learning contexts.

Some proponents of authentic classroom language models have argued that the problems with many textbooks are not due to the fact that they are culturally or socially biased but that they are actually too contrived and artificial in their presentation of the target language. They argue that it is crucial to introduce learners to the fundamental characteristics

of authentic real life examples of both spoken and written discourse. They have demonstrated that many scripted textbook language models and dialogues are unnatural and inappropriate for communicative or cooperative language teaching because they do not adequately prepare students for the types of pronunciation (Brazil, Coulthard, & Johns, 1980; Levis, 1999), language structures, grammar, idioms, vocabulary and conversational rules, routines and strategies that they will have to use in the real-world (Cathcart, 1989; Bardovi-Harlig, Hartford, Mahan-Taylor, Morgan, & Reynolds, 1991; Yule, Matthis, & Hopkins, 1992). They further contend that the scripted unauthentic language found in many textbooks does not lend itself to communicative practice but instead can lead to an oversimplification of language and unrealistic views of real-life situations. It can also provide additional inaccurate advice about the target language society that can be particularly dangerous for students entering the target language community or those who are expecting to engage in significant amounts of real-life interactions with native speakers.

A final reason for disappointment and skepticism with many ELT textbooks is the fact that they are often regarded as the "…tainted end-product of an author's or a publisher's desire for quick profit" (Sheldon, 1988, p. 239). Too many textbooks are often marketed with grand artificial claims by their authors and publishers, and yet these same books tend to contain serious theoretical problems, design flaws, and practical shortcomings.

1.1 Justification for Textbook Evaluation

Since the 1970's, there has been a movement to make learners the center of language instruction. It is probably best to view textbooks as resources in achieving aims and objectives that have already been set in terms of learner needs. Moreover, they should not necessarily determine the aims themselves (components of teaching and learning) or become the aims but they should always be at the service of the teachers and learners (Brown, 1995).

Consequently, we must make every effort to establish and apply a wide variety of relevant and contextually appropriate criteria for the evaluation of the textbooks that we use in our language classrooms. We should also ensure "that careful selection is made, and that the materials selected closely reflect [the needs of the learners and] the aims, methods, and values of the teaching program" (Cunningsworth, 1995, p.7).

Sheldon (1988) offers several other reasons for textbook evaluation. He suggests that the selection of an ELT textbook often signals an important administrative and educational decision in which there is considerable professional, financial, or even political investment. A thorough evaluation, therefore, would enable the managerial and teaching staff of a specific institution or organization to discriminate between all of the available textbooks on the market. Moreover, it would provide a sense of familiarity with a book content thus assisting educators in identifying the particular strengths and weaknesses in textbooks already in use. This would go a long way in ultimately assisting teachers with making optimal use of a book's strong points and recognizing the shortcomings of certain exercises, tasks, and entire texts.

One additional reason for textbook evaluation is the fact that it can be very useful in teacher development and professional growth. Cunningsworth (1995) and Ellis (1997) suggest that textbook evaluation helps teachers move beyond impressionistic assessments and helps them to acquire useful, accurate, systematic, and contextual insights into the overall nature of textbook material. Textbook evaluation, therefore, can potentially be a particularly worthwhile means of conducting action research as well as a form of professional empowerment and improvement. Similarly, textbook evaluation can also be a valuable component of teacher training programs for it serves the dual purpose of making student teachers aware of important features to look for in textbooks while familiarizing them with a wide range of published language instruction materials. As such, the present study focuses on an issue (evaluation of SAMT ESP textbooks). Many teachers and students have commented

on the inappropriateness and shortcomings of the SAMT textbooks, but they do not have the time and the authority to explore and publicize the issue.

2. Literature Review

The present section is a brief review of the studies conducted on textbook evaluation. As such, the introductory part of the section begins with presenting some overseas studies textbook evaluation. The second and the third parts of the section address the studies conducted on general and ESP textbook evaluation in Iran.

2.1 Overseas Studies on textbook evaluation

The first group of textbook studies focused on determining the readability levels of curriculum materials. The readability studies reviewed demonstrated that content area textbooks such as social studies and sciences are written at reading levels several years above the reading ability of the intended audience (Ford, 1972; Clarke, 1977; Campbell, 1972; Santa & Burstyn, 1977; Marshall, 1979;

A second category of studies on textbook evaluation utilize Bloom's (1974) Taxonomy of Educational Objectives as their criteria. Their purpose was to determine the level of thought processes emphasized in textbooks. Findings from investigations conducted by Reynolds (1967), Schmidt (1971), Withrow (1971) and Nicholson (1977) revealed the idea that the analyzed textbooks emphasize primarily the knowledge level of Bloom's taxonomy.

A final area of textbook evaluation research includes studies that deal with the teaching methods and suggestions provided in manuals or textbooks. These studies generally determine the kind of instruction provided, the sequence of instruction and learning activities, and the amount of emphasis or page space devoted to instruction. These elements are then

compared to the emphasis placed on assessment, assignment, theory and practice activities. Therefore, some of the relevant studies related to the last category are presented below.

Chadran and Essarey (1997) conducted a study about English textbooks used in Malaysian schools. The results show that, in general, teachers prefer commercially produced materials available in the market over the prescribed textbooks developed by the Ministry, that they do not engage themselves in producing materials of their own, that they consider the textbooks out-dated and dull, and that textbooks are not suitably graded in terms of difficulty.

Kim (2001) develops his own Communicative Language Teaching (CLT) criteria to study the 6th and 7th grade national Korean English curricula. The study reveals that despite the recent attempt to implement CLT-based elementary EFL, there are still deficiencies in the Korean elementary EFL curriculum and materials, such as the use of the Audio-lingual approach in teaching the 7th curriculum material set.

Tomlinson (2001) mentions two factors contributing to a change in attitude toward materials development which makes it an independent field of its own. One is the realization that one fruitful way to familiarize teachers with professional aspects of language teaching is through helping them to engage in the process of materials development and their evaluation. The other is the realization that no single textbook can possibly be ideal for all situations and learners, and hence teachers need to evaluate and adapt the materials they use in classroom and to evaluate the materials of their own in order to guarantee a match between the students' variables and the materials.

Melo's (2003) study is an evaluation and analysis of a textbook series called True Colors (Levels 1, 2, 3, and 4) and their respective workbooks used to teach English in Brazil. The researcher concludes that teachers who use those textbooks do not seem to be satisfied with them although there is some positive feedback from some of the participants.

In the realm of textbook evaluation, Otlowski (2003) investigates the portrayal of gender and the representation of the various ethnic groups in the Expressway A series. The results show that Expressway A, while better than many earlier EFL texts, still depicts women in roles that no longer accurately represent their role in society.

The findings of Vellenga's (2004) study show that the textbooks include a paucity of meta-linguistic and explicit meta-pragmatic information, and the comparison of EFL and ESL textbooks reveals that although the amount of pragmatic information is small across all texts, a larger percentage of pages of EFL texts are comprised to pragmatic information; however, the quality of pragmatic information is better in terms of number of speech acts presented and amount of meta-pragmatic cues in ESL texts.

In his doctoral dissertation, Madkhali (2005) proposes an alternate syllabus design for the one currently practiced at the Institute of Public Administration (IPA) in Saudi Arabia, Riyadh. The goals for this new design are based on the findings of needs analyses at two settings: English for Academic Purposes (EAP) at the IPA, and English for job purposes at the Saudi private sector department (PSD). The results show that EAP students face difficulties in three English language skills: reading, writing, and speaking. Similarly, the perceptions of PSD students were that to a certain degree they face difficulties in the four language skills. Based on the needs found, the researcher suggests teaching two business courses: Business Correspondence and Business Communication. In addition, the researcher proposes an alternate design for the current program at the IPA's English Language Center which should focus on General English, rather than English for Specific Purposes.

Al-Saif (2005) conducted a comprehensive evaluation of the textbook English for Saudi Arabia which is taught in the sixth grade in Saudi elementary schools. The findings of the study showed that the distribution of data was not conclusive. It summarized that the textbook was only "moderately adequate", and that it required constant and cyclical

evaluation based on the changing circumstances of the program. Al-Saif (2005) argues that this finding "suggests some indecisiveness on the part of the evaluators" (p. 62), caused by their limited training and their experience of the textbook. This shows that in addition to the level of curriculum implementation, teacher professional development is a significant factor in any successful textbook evaluation.

2.2 Studies on general textbook evaluation in Iran

Amerian (1987) used Tucker's (1975) system of textbook evaluation to analyze two series of textbooks: 'Right Path to English' and 'Graded English'. His analysis showed that the two were almost the same in that they were both 'structural' in nature of syllabus design. As he believes, Tucker's scheme was most suitable for the job in that it was also structural and examined sub-systems of language (pronunciation, grammar, vocabulary, etc.) separately.

Kheibari (1999) modified Tucker's model and applied it to the five volumes of teaching Persian to speakers of other languages (TPSOL) text books. She claimed that the philosophy behind the changes, due to the recent developments in language teaching the results revealed that the books follow the grammar translation method which attaches the least attention to role-playing, and different kinds of tasks or language skills such as speaking.

Shahedi (2001) analyzed one of the leading texts in Teaching Persian to Speakers of Other Languages (TPSOL) and stated that in these series, not enough attention has been attached to the four skills of the language. Moreover, the manner and amount of the presentation of vocabulary and pronunciation are not in harmony with language learners' proficiency levels.

Ansary and Babaii (2002) analyzed a corpus of 10 EFL/ESL textbook reviews plus 10 EFL/ESL textbook evaluation checklists and outlined what is perceived to be the common core features of standard EFL/ESL textbooks. The major categories comprise approach,

content presentation, physical make-up and concerns. Each set of major features of EFL/ESL textbooks consists of a number of subcategories. The researchers concluded that not all of these characteristics are dominant in all textbooks.

Yarmohmmadi (2002) evaluated the senior high school textbooks based on a revised version of Tuckers' model. He came to the conclusion that these textbooks suffer from a lot of shortcomings: 1. they are not authentic; 2. English and Persian names are used interchangeably; and 3. oral skills are ignored. At the end, some suggestions were proposed to remedy the shortcomings.

Iraji (2007) conducted a study and made a careful analysis on New Interchange series based on the principles of communicative and task-based approach to investigate to what extent the principles of CLT and TBLT approaches have been regarded. In this regard, she employs Ellis's (2003) model. Iraji (2007) criticizes New Interchange because the series do not follow the principles of communicative and task-based approaches as the author has claimed. It does not have frequency of meta-pragmatic information. The distributional pattern of communicative activities were random and without pattern.

Razmjoo (2007) investigated the extent to which the Iranian high school and private institute textbooks represent the CLT principles. To this end, the textbooks of the Iranian high schools and private institutes were analyzed descriptively and inferentially. The analysis of the data indicated that while high school textbooks are not conductive to CLT implementation, private institute textbooks represent the CLT principles to a great extent.

2.3 Studies on ESP textbook evaluation in Iran

Sadeghian (2005) called for a reappraisal of the philosophy and pragmatics of ESP as well as the role of ESP materials and teacher, and mentioned that the main purpose of any ESP

purpose program should be the development of an appropriate culture, rather than a special language.

Eslami Rasekh (2005) highlighted the importance of pragmatically hidden assumptions in ESP books. She discussed Speech Acts, Cooperative Principle, and Politeness Theory to show that ESP texts are socially constrained and emphasizes that ESP teacher should be equipped with the knowledge and use of pragmatics.

Maleki's (2005) study focused on answering the controversial question of whose job it is to teach ESP. Citing the paradigms shift in ESP over the past three decades, beginning from register analysis, he found a theoretical support for the claim that the language teacher is the one solely qualified for teaching ESP classes.

Tayebipour (2005) addressed some major inadequacies of ESP curriculum in Iran in terms of instructors' competence and the design of the materials. He criticized the current read-only approach for being too narrow and inefficient, and offered some suggestions to improve the practice of ESP/EAP in the country.

Manafi Anari (2005) dealt with what he considers to be the main drawbacks of ESP practice in Iran, in his paper. He identified three major problems: un-preparedness of the majority of students for their ESP courses, the ESP textbooks' unduly high level of difficulty, and inappropriate materials design and development and at last he suggested certain ways of overcoming these problems.

Sadeghi (2005) suggested a carryover from the existing situation in his paper. He considered ESP in terms of student, teacher, materials, and methodology, He highlighted the inadequate language skills of students upon entering university, and most ESP teachers' incompetence with respect to either language or subject knowledge. He attempted to highlight the importance of communicative and learning-centered approaches to ESP practice and material production.

Farhady (2005) discussed ESP parameters, namely needs, materials, method, learner, teacher, and context. He began with a brief history of materials development SAMT and concluded by a series of suggestions to improve ESP in Iran, such as carrying out thorough need analyses, changing the design and concepts of the materials, training qualified teachers, and reforming the testing procedures.

Mozayan (2005) elaborated on the effect of schematic mapping on the improvement of EAP reading comprehension ability among the Iranian students of Medical Sciences. He emphasized the use of diagrams (semantic mapping) to reinforce text comprehension and lent empirical support to his argument through a pilot study.

Tajeddin (2005) evaluated some of the ESP textbooks regarding three aspects, namely, linguistic input, linguistic output and their relationship to each other with regard to the applied methodology based on three approaches of bottom-up, top-down and interactional processing. He concluded that ESP textbooks published by SAMT organization not only do lack one integrated approach and lesson plan but also do not follow any specific purpose regarding the selected reading passages, or activities designed for improving the learners' translation or comprehension ability, at all. He then suggested that the future planning requires a fundamental and basic revision of the textbooks.

2.4 The contribution of the Literature Review to the Present Study

Having reviewed the studies done on the textbook evaluation in Iran and the other countries, the researchers have come to the conclusion that textbooks play a very crucial role in language teaching and learning, and it is the next important factor in second or foreign language classrooms after the teacher. Moreover, the present study is supported by Riazi's (2003) ideas regarding the importance of textbook evaluation:

1. Textbook evaluation allows us to select a newly started language program.

2. Evaluation of books already in use in language program is very important to identify the strengths and weaknesses of the books.
3. We can evaluate books after a course of instruction with the objective of retaining, updating or substituting the books.

Considering these points and because of the importance of textbook evaluation, the researchers intended to investigate the weaknesses and strengths of these books to inform SAMT organization and hoped to propose some practical suggestions to remedy the shortcomings.

2.5 The Objective and significance of the study

"Tell me what you need English for and I will tell you the English that you need". This is the guiding principle of ESP as stated by Hutchinson and Waters (1987, p.8). The need of a particular group of learners could be identified by the linguistic characteristics of their specialist area of work or study.

The overall objective of the present study is to identify the organizational patterns of the SAMT ESP textbooks and also to evaluate whether they meet the learners' needs or not. In other words, the purpose of this study is to evaluate ESP textbook materials regarding various internal and external aspects and characteristics of those materials in order to give an overall view of their content and method. Specifically, this study intends to achieve three objectives:

1. Evaluating SAMT ESP textbooks designed for the students of Medical Sciences from the students' point of view
2. Evaluating SAMT ESP textbooks designed for the students of Medical Sciences from the instructors' point of view
3. Comparison and contrast of the students' and instructors' viewpoints

Regarding the significance of the study, the results of this study would be useful for three groups, namely:

1. ESP material developers and designers
2. Students and instructors at Medical Sciences universities; and
3. Finally, SAMT organization

2.6 Research questions

Based on the objectives of the study, the following questions are addressed in this study:

1. Are SAMT textbooks designed for the students of Medical Sciences appropriate from students' and instructors' viewpoints in terms of theoretical construction, the organizational features and practical considerations from students' and instructors' viewpoints, the content, language skills, the vocabulary, and the structure?
2. What is the students' and instructors' overall impression of the textbooks?
3. Is there any significant difference among the students' and instructors' viewpoints?

3. Methods of The study

This section introduces the participants and materials, specifying how, where and in what ways they were selected. Moreover, the instruments used for data collection will be presented. In addition, the data analyses along with the procedures made use of will be elucidated.

3.1 Participants

As this study was designed to evaluate ESP textbooks for the students of medical Sciences published by SAMT organization, the researchers decided to seek ESP instructors and students' viewpoints at different Iranian universities, namely, Shiraz University of Medical Sciences, Islamic Azad University and Iran University in Tehran benefiting from the

availability sampling. Therefore, 43 questionnaires were distributed among the above-mentioned university instructors. Due to the fact that not all those instructors cooperated with the researchers, the number of participants was reduced to only 30 instructors whose viewpoints and suggestions were gathered through the questionnaires. All participates were highly experienced in the fields of general and academic English language teaching.

As the next step, the researchers distributed the questionnaires among 147 students who studied ESP in Medical Sciences at the same universities because they studied these courses and had enough experience concerning the strengths and weaknesses of these kinds of textbooks as learners. Their views and suggestions could be considered as a rich source of data to be handled during the research procedure.

3.2 Materials

The material to be evaluated is a sample of the ESP textbooks published for the students of Medical Sciences by SAMT. This organization has published about 56 titles of ESP books and almost all of the Iranian universities practice the SAMT textbooks. There are about 9 books being taught to the university students in Medical Sciences consisting of: English for the students of Occupational health and Safety; English for the students of Laboratory Technology; English for the students of Pharmacy; English for the students of Nursing; English for the students of Veterinary Medicine; English for the students of Dentistry; English for the students of Medicine I; English for the students of Medicine II; and English for the students of Medicine III.

3.3 Instruments

The instrument in this study is a questionnaire with a likert scale format developed by Tahriri and Shahini (personal communication) which was originally based on Littlejohn's (1998)

scheme. Since their textbook evaluation form has been designed for General English textbooks, some adaptation was deemed necessary to make it more appropriate for ESP materials. There are two columns in the questionnaire for each main part related to the Necessity and Present Conditions respectively. The first column which is indicative of necessity condition ranges from 1 to 4 representing "not necessary" to "highly necessary" while the second column introduces the present condition ranging from 1 to 5 which shows "totally lacking" to "excellent" (See Appendix A). It should be added that for the sake of easy and informative comparison and contrast of the two conditions, 'good' and ' excellent' options have been merged in the present condition. Accordingly, 'very good' option has been used instead.

In fact, the questionnaire includes some general overview of physical aspects of the material such as the organizational features and practical considerations with respect to cover, size, or durability of the book which corresponds to the "publication" in Littlejohn's (1998) scheme while the other part relates to the thinking underlying the materials such as the theoretical considerations focusing on the methodological objectives of the textbook or the sequencing of the selected subjects, types of learning and teaching activities with respect to content, skills, vocabulary, and structure of the book which corresponds to the "design" in Littlejohn's (1998) scheme.

3.4 Validity and Reliability of the Questionnaire

Factor analysis was run on both sections of the scales of the questionnaire, including the necessity and the present conditions. The results for both parts show that there are six factors or components loading on each part of the questionnaire. As for reliability, Chronbach alpha was run and the obtained index is 0.92. Experts in statistics believe a minimum of 0.70 is a fairly reasonable index for an instrument to be reliable. This is indicative that the instrument

used in the study is appropriate in terms of reliability. Alagha (2007) followed the same procedures and reported almost the same indices of reliability and validity.

3.5 Data Collection and Analysis Procedure

In the first phase, the questionnaire was distributed among the ESP instructors individually in their office hours. Their viewpoints and suggestions about ESP textbooks designed for the 9 different courses of ESP were collected. In the second phase, the same questionnaire was distributed among the ESP students to collect their viewpoints and suggestions according to their needs concerning the textbooks designed for different fields of "Medicine". All the data gathered through questionnaires were analyzed both descriptively (Mean + Standard Deviation + Graph) and inferentially (paired and sample t-tests) as presented in the next session.

4. Results and Discussion

This section presents the results of the analyses of the participants' responses to the survey. As such, first the instructors' ideas are analyzed and presented; however, for the sake of brevity, the results of the students' ideas are presented in Appendix B.

Criterion 1: Instructors' views on ESP Textbooks: Theoretical Considerations (Instructors)

Table 1. Descriptive Statistics for the Theoretical Considerations

Necessity Condition	Not necessary **(NN)**		Almost necessary **(AN)**		Necessary **(N)**		Highly necessary **(HN)**	
	***F**	****P**	**F**	**P**	**F**	**P**	**F**	**P**
	4	**13.3**	**5**	**16.7**	**8**	**26.7**	**13**	**43.3**
Present Condition	Totally Lacking **(TL)**		Poor (**P**)		Adequate **(A)**		Very Good (**VG**)	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
	12	**40**	**9**	**30**	**6**	**20**	**3**	**10**

*F=Frequency **P=Percentage

As shown in this table, 13.3% of the participants rated the theoretical considerations of the ESP textbooks as 'not necessary' for them, 16.7% as 'almost necessary', 26.7% as 'necessary', and 43.3% as 'highly necessary'; while for the present condition, 40% of the participants have selected 'totally lacking', 30% 'poor', 20% 'adequate', and 10% 'very good' choices.

Table 2. Theoretical consideration: Paired-sample t-test between 'necessity' and 'present' conditions

Theoretical consideration	N	Mean	S. D	t	Sig.
Present	30	2	1.01710	3.144	.004*
Necessity	30	3	1.08278		

* Significant at .05

According to Table 2, the difference between 'necessity' and 'present' conditions of the textbooks is statistically significant. This means that the difference between the ideal situation and the present condition of the textbooks is real and the textbooks are lacking in terms of setting goals and objectives.

Criterion 2: Organizational Features and Practical Considerations (Instructors)

Table 3. Descriptive Statistics for the Organizational Features and Practical Considerations

Necessity	**NN**		**AN**		**N**		**HN**	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**5**	**16.7**	**6**	**20**	**10**	**33.3**	**9**	**30**
Present	**TL**		**P**		**A**		**VG**	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**7**	**23.3**	**11**	**36.7**	**5**	**16.7**	**7**	**23.3**

As indicated above, 16.7% of the participants have selected 'not necessary', 20% 'almost necessary', 33.3% 'necessary', and 30% 'highly necessary' options for the necessity condition. Moreover, it is clear that 23.3% of the participants have opted for 'totally lacking', 36.7% 'poor', 16.7% 'adequate', and 23.3% 'very good' options.

Table 4. Organizational Features and Practical Considerations: Paired-sample t-test between 'necessity' and 'presents conditions

Organizational Features	N	Mean	S.D	t	Sig.
Present	30	2.4	1.10172	-4.097	.000*
Necessity	30	2.7667	1.07265		

* Significant at .05

According to Table 4, the difference between 'necessity' and 'present condition' of the textbooks is statistically significant. This means that the difference between the ideal situation and the present condition of the textbooks is real and they are not appropriate in terms of appealing cover, appropriate size, durability, clear layout, useful table of content, glossary, and index, accurate appendices, references and resources, logical arrangement, clear and accurate diagrams, figures and pictures.

Criterion 3: Content (Instructors)

Table 5. Descriptive Statistics for the Content

Necessity	NN		AN		N		HN	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**3**	**10**	**6**	**20**	**12**	**40**	**9**	**30**
Present	TL		P		A		VG	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**7**	**23.3**	**14**	**46.7**	**4**	**13.3**	**5**	**16.7**

As the table shows, 10% of the participants have chosen 'not necessary', 20% 'almost necessary', 40% 'necessary', and 30% 'highly necessary' options. In addition, 23.3% of the participants have selected 'totally lacking', 46.7% 'poor', 13.3% 'adequate', and 16.7% 'very good', and 30% 'highly necessary' options for the present condition.

Table 6.Content: Paired-sample t-test between 'necessity' and 'present' conditions

Content	N	Mean	S.D	t	Sig.
Present	30	2.2333	1.00630	2.307	.028*
Necessity	30	2.9	.95953		

* Significant at .05

According to Table 6, the difference between 'necessity' and 'present condition' of the books is statistically significant. This means that the difference between the ideal situation and the present condition of the books is meaningful and the books lack the relevance of the content to the aims of the curriculum and syllabub, topic covering of ESP texts, up-to-date content, logical organization of the subject manner, authentic texts, interesting content, appropriate activities, and correspondence between the content and student's discipline.

Criterion 4: Skills (Instructors)

Table 7. Descriptive Statistics for the Skills

Necessity	NN		AN		N		HN	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**3**	**10**	**4**	**13.3**	**10**	**33.3**	**13**	**43.3**
Present	TL		P		A		VG	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**6**	**20**	**12**	**40**	**7**	**23.3**	**5**	**16.7**

As indicated in Table 7, 10% of the participants have opted for 'not necessary', 13.3% 'almost necessary', 33.3% 'necessary', and 43.3% 'highly necessary' choices; whereas 20% of the participants have selected 'totally lacking', 40% 'poor', 23.3% 'adequate', and 16.7% 'very good' choices for the present condition.

Table 8. Skills: Paired-sample t-test between 'necessity' and 'present' conditions

Skills	N	Mean	S.D	t	Sig.
Present	30	2.3667	.99943	2.947	.006*
Necessity	30	3.1	.99481		

* Significant at .05

Once again the difference is clearly significant suggesting that the activities in the textbook are not balanced among various skills, an adequate number of skill-building activities are not

provided, each skill is not treated in every unit, no balance is observed among the various skill-building activities, a mismatch is found between the activities and the required skills, skill-building strategies are not included in the texts, no logical development of the skills throughout the textbook is noticed and the activities activating student's background knowledge are not included.

Criterion 5: Vocabulary (Instructors)

Table 9. Descriptive Statistics for the Vocabulary

Necessity	NN		AN		N		HN	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**2**	**6.7**	**5**	**16.7**	**7**	**23.3**	**16**	**53.3**
Present	TL		P		A		VG	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**2**	**6.7**	**4**	**13.3**	**7**	**23.3**	**17**	**56.7**

According to Table 9, for the necessity condition, 6.7% of the participants have chosen 'not necessary', 16.7% 'almost necessary', 23.3% 'necessary', and 53.3% 'highly necessary 'options. Moreover, 6.7% of the participants have opted for 'totally lacking', 13.3% 'poor', 23.3% 'adequate', and 56.7% 'good' choices.

Table 10. Vocabulary: Paired-sample t-test between 'necessity' and 'present' conditions

Vocabulary	N	Mean	Std. Deviation	t	Sig.
Present	30	3.3	.95231	-.290	.774
Necessity	30	3.2333	.97143		

As shown in Table 10, there is no significant difference between the ideal situation and present condition of the vocabulary used in the ESP textbooks; that is, the vocabularies used in these textbooks are appropriate for ESP students from the instructor's viewpoints.

Criterion 6: Structure (Instructors)

Table 11. Descriptive Statistics for the Structure

Necessity	NN		AN		N		HN	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**3**	**10**	**4**	**13.3**	**8**	**26.7**	**15**	**50**
Present	TL		P		A		VG	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**2**	**6.7**	**6**	**20**	**4**	**13.3**	**18**	**60**

According to Table 11, 10% of the participants have opted for 'not necessary', 13.3% 'almost necessary', 26.7% 'necessary', and 50% 'highly necessary' options for the necessity condition while, 6.7% of the participants have chosen 'totally lacking', 20% 'poor', 13.3% 'adequate', and 60% 'very good' choices for the present condition.

Table 12. Structure: Paired-sample t-test between 'necessity' and 'present' conditions

Structure	N	Mean	S.D	t	Sig.
Present	30	3.2667	1.01483	-.351	.728
Necessity	30	3.1667	1.01992		

As stated in Table 12, there is no significant difference between the present condition and ideal situation of the Medicine ESP textbooks regarding structure.

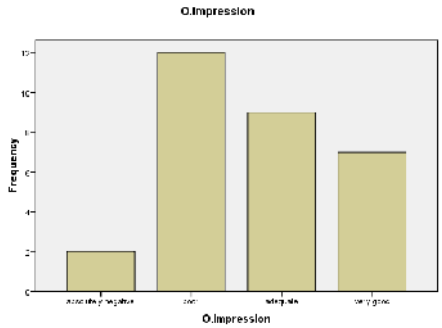
Overall Impression (Instructors)

Table 13. Descriptive Statistics for the Overall impression (Instructors)

	TL		P		A		VG	
Principles	**F**	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Overall Impression	**2**	**6.7**	**12**	**40**	**9**	**30**	**7**	**23.3**

Table 13 indicates that 6.7% of the participants have opted for 'totally lacking', 40% 'poor', 30% 'adequate', and 23.3% 'very good' choices. The schematic presentation of the data is presented in Figure 1.

Figure 1. Display of the Data for Overall impression (Instructors)



Overall analysis of the Instructors' Viewpoints on the "Present" and "Necessity" Conditions of ESP textbooks

Table 14. Paired Sample t-test among 30 instructors' viewpoints regarding ESP textbooks

Instructors' Viewpoints	N	Mean	S.D	t	Sig.
Present	30	2.6	1.25782	3.455	.001*
Necessity	30	3.0278	1.01624		

* Significant at .05

A paired sample t-test was run between instructors' views on the necessity and the present conditions concerning the six main parts of the questionnaires. The results showed a significant difference (0.00) at $p \leq 0.05$ between two conditions. These outcomes prove that there is a significant difference between the ideal situation and the present condition of the evaluated textbooks. As such, they may be placed in a position that is far from the standard or the ideal one. As mentioned before, for the sake of brevity, the results of the students' ideas are presented in Appendix B.

Comparison of Students' and Instructors' Viewpoints Regarding ESP textbooks

Table 15. Independent Sample t-test between instructors' and students' viewpoints

Conditions	**Participants**	N	Mean	SD	t	Sig.
Necessity Condition	Students	147	2.9342	1.04071	1.103	.163
	Instructors	30	3.0278	1.01624		
Present Condition	Students	147	2.6757	1.11702	-.811	.002*
	Instructors	30	2.6000	1.25782		

* Significant at .05

According to Table 15, there is no significant difference between instructors' and students' viewpoints with respect to the necessity condition of these ESP textbooks whereas there is a significant difference between instructors' and students' viewpoints regarding the present condition of these ESP textbooks.

5. Conclusion

A language course and textbook should help students communicate effectively in their everyday encounter, develop all four skills, enhance their ability to solve problems, understand the general features of the language, discuss, ask and answer questions related to current topics, write and present summaries, enhance their vocabulary.

To sum up the study, the main points can be encapsulated by answering the research questions.

1. Are the textbooks designed by SAMT publication for the students of Medicine appropriate in terms of theoretical considerations from students' and instructors' viewpoints?

Neither the instructors nor students were satisfied with the theoretical considerations of ESP Medicine textbooks.

2. Are the textbook designed by SAMT publication for the students of Medicine appropriate in terms of organizational features and practical considerations from students' and instructors' viewpoints?

The instructors were not satisfied with the organizational features and practical considerations of the ESP Medicine textbooks, but the students found them satisfactory.

3. Are the textbooks designed by SAMT publication for the students of Medicine appropriate in terms of content from students' and instructors' viewpoints?

The content of ESP Medicine textbooks was not found to be satisfactory by both the instructors and students.

4. Are the textbooks designed by SAMT publication for the students of Medicine appropriate in terms of skills from students' and instructors' viewpoints?

The instructors were not satisfied with the ESP Medicine textbooks in terms of the skills, but it was not the case for students.

5. Are the textbook designed by SAMT publication for the students of Medicine appropriate in terms of vocabulary from students' and instructors' viewpoints?

The instructors were satisfied with the ESP Medicine text books in terms of vocabulary, but the students were not.

6. Are the textbooks designed by SAMT publication for the students' of Medicine appropriate in terms of structure from students' and instructors' viewpoints?

The structure of the ESP Medicine books was satisfactory for the instructors, but not for the students.

7. What is the students' and instructors' overall impression of the textbooks?

Both the instructors and students were not satisfied with the content of ESP Medicine textbooks

8. Is there any significant difference among students' and instructors' viewpoints?

No significant difference was found between the students' and instructors' opinions regarding the necessity condition of ESP Medicine textbooks, but a significant difference was found

between the students' and instructors' opinions with regard to the present condition of ESP Medicine textbooks.

6. Pedagogical Implications

This section of the study aims at illustrating and incorporating the theoretical findings of the attempt in practical areas such as second or foreign language learning and teaching, material development, and test construction. The findings of this study can be both theoretically and practically significant for the following reasons: Iranian teachers, especially, ESP ones are expected to benefit from this analysis in order to conceptualize a crystal clear idea about the nature of the textbooks they teach. This study can also provide material developers and textbook writers with the necessary information one needs for a fully-fledged textbook. Textbook designers could mind the pitfalls of the textbook under analysis to develop and design a new textbook. Textbook designers and material developers should provide learners with more real life exercises which pave the way to involve both teachers and learners in classroom activities as much as possible. One possible way is including more different kinds of tasks like ‘problem-solving tasks’, ‘guessing-tasks’, and ‘opinion-tasks’, which need real communication for enthusiastic students to take part in classroom activities. What matters in these activities is that students should have a desire to communicate in the target language. Specific activities should be provided for the learners to communicate (e.g. to make a point, to buy an airline ticket, or write a letter to a newspaper). They should focus on the content of what they are saying or writing rather than on a particular language form. They should use a variety of language forms rather than just one language structure. In other words, activities should attempt to replicate real communication.

Yet, despite the fact that there are some limitations to this study, it has been the first comprehensive study on ESP Medicine textbooks to evaluate them from a variety of aspects in order to provide an objective and complete analysis of their strengths and weaknesses.

Acknowledgments

We are grateful to the anonymous reviewers of the study for their care, time and meticulous comments. We would also like to express our special thanks to Prof. Damian Rivers and Prof. John Adamson, the Senior Editors of the Asian ESP Journal, for their contribution.

References

Alagha, N. R. (2007). An Evaluation of a sample of ESP textbooks for Humanities published by SAMT. Unpublished master's thesis, Shiraz Universiy, Shiraz.

Allwright, R. L. (1981). What do we want teaching materials for? *ELT Journal, 36*(1), 5–18.

Alptekin, C. (1993). Target-Language Culture in EFL Materials. *ELT Journal. 47*(2), 136-143.

Al-Saif, A. (2005). An evaluation of the TEFL textbook for the 6th elementary class in Saudi Arabia. Unpublished master's thesis, Essex University, UK.

Amerian, M. (1987). A comparative study of the graded English and the right path to English series with regard to content and methodology. Unpublished master's thesis, Shiraz University, Shiraz.

Ansary, H., & Babaei, E. (2002). Universal characteristics of EFL/ESL textbooks: A step towards systematic textbook evaluation. *The Internet TESL Journal, 2,* 1-8.

Bardovi-Harlig, K., Hartford, B. A. S., Mahan-Taylor, R., Morgan, M. J., & Reynolds, D. W. (1991): Developing pragmatic awareness: Closing the conversation. *ELT Journal, 45*(1), 4-15.

Bloom, B. (1994). Reflections on the development and use of taxonomy. In Anderson, L. & Sosniak, L. (Ed.), *Bloom's Taxonomy: A Forty-Year Retrospective* (pp. 1-8). Chicago: The National Society for The Study of Education.

Brazil, D., Coulthard, M., & Johns, C. (1980). *Discourse Intonation and Language Teaching*. Harlow: Longman

Brown, J. (1995). *The Elements of Language Curriculum*. Boston, MA: Heinle and Heinle Publishers.

Campbell, B. R. (1972). A study of the relationship of reading ability of students in grades 4, 5, and 6, and comprehension of social studies and science textbook Selections. Unpublished doctoral dissertation, University of Florida, United States.

Carrell, D., & Korwitz, J. (1994). Using Concordancing Techniques to Study Gender Stereotyping in ELT Textbooks. In J. Sunderland (Ed.), *Exploring Gender: Questions and Implications for English Language Education* (pp. 31-44). Prentice Hall International.

Cathcart, R. (1989). Authentic Discourse and the Survival English Curriculum. *TESOL Quarterly, 23*(1), 105-126.

Chadran, J. & Essarey, G. (1997). Content-based instruction: An Indonesian example. In S. B. Stryker and B. L. Leaver (Ed.), *Content-based Instruction in Foreign Language Education* (pp. 219-235). Washington D.C.: Georgetown University.

Clarke, B. K. (1977). Eighth grade students' reading abilities and their comprehension of selected social studies and science textbooks. Unpublished doctoral dissertation, University of Florida, United States.

Clarke, J., & Clark, M. (1990). Stereotyping in TESOL materials. In B. Harrison (Ed.), *Culture and the language classroom* (pp. 31-44). Hong Kong: Modern English Publications and the British Council.

Cunningsworth, A. (1995). *Choosing your coursebook*. Oxford: Heinemann.

Ellis, R. (1997). The empirical evaluation of language teaching materials. *ELT Journal, 51(1)*, 36-41.

Ellis, R. (2003). *Task based Language Learning & Teaching*. Oxford: Oxford University Press.

Eslami Rasekh, Z. (2005). The Relevance and Sociocultural Significance of Pragmatics for ESAP. Paper presented in the First National ESP/EAP Conference, Tehran, Iran.

Farhady, H. (2005). 25 Reflections on and Directions for ESP Materials Development in SAMT. Paper presented in the First National ESP/EAP Conference, Tehran, Iran.

Florent, J., & Walter, C. (1989). A Better Role for Women in TEFL. *ELT Journal. 43*(3),180-184.

Ford, L. (1972). A descriptive study of the readability grade level of the most frequently state adopted intermediate level social studies textbooks. Unpublished doctoral dissertation, University of McNeese State, United States.

Gray, J. (2000). The ELT Coursebook as Cultural Artifact: How teachers censor and adapt. *ELT Journal, 54*(3), 274-283.

Haycroft, J. (1998). *An Introduction to English Language Teaching*. London: Longman.

Hutchinson, T., & Torres, E. (1994). The textbook as agent of change. *ELT Journal, 48(4),* 315-327.

Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning centered approach.* Cambridge: Cambridge University Press.

Iraji, R. (2007). Does new interchange books reflect task-based or communicative approaches. *ILI Language teaching Journal, 4*, 25-38.

Kheibari, S. (1999). Text analysis and evaluation of TPSOL course books. Unpublished master's thesis, Shiraz Universiy, Shiraz.

Kim, H. J. (2001). A case study of curriculum and material evaluation: Elementary English as a foreign language in South Korea. Unpublished master's thesis, McGill University, Canada.

Kitao, K., & Kitao, S. K. (1997). Selecting and Developing Teaching/Learning Materials? *The Internet TESL Journal, IV* (4) http://iteslj.org/Articles/Kitao-Materials.html.

Levis, J. (1999). Intonation in Theory and Practice Revisited. *TESOL Quarterly, 33*(1), 37-63.

Littlejohn, A. (1998). The analysis of language teaching materials: Inside the Trojan Horse. In B. Tomlinson (Ed.), *Materials development in language teaching* (pp. 190-216). Cambridge: CUP.

Mabdkhali, H. M. (2005). A language curriculum model: A case study in Saudi Arabia. Unpublished doctoral dissertation. Ball State University, Muncie Indiana.

Maleki, A. (2005). Medicine or Medical? Paper presented in the First National ESP/EAP Conference, Tehran, Iran.

Manafi Anari, S. (2005). Problems of achieving to ESP purposes in the universities. Paper presented in the First National ESP/EAP Conference, Tehran, Iran.

Marshall, N. (1979). Readability and comprehensibility. *Journal of Reading, 22(6),* 542-544.

Melo, R. (2003). Teachers' opinions about True color series textbooks. *International Journal of Educational Technology, 3*, 78-86.

Mozayan, M. R. (2005). A Sample Lesson Proposal. Paper presented in the First National

Nicholson, J. I. (1977). An analysis of the levels of comprehension required by questions in teachers' manuals of selected basic reading textbooks. Unpublished doctoral dissertation, University of Mississippi State, United States.

O'Neill, R. (1982). Why use textbooks? *ELT Journal, 32(2),* 148-156.

Otlowski, M. (2003). Ethnic diversity and Gender bias in EFL textbooks. *Asian EFL Journal, 5(3)*, 18-27.

Phillipson, R. (1992). *Linguistic Imperialism*. Oxford: Oxford University Press.

Porreca, K. L. (1984). Sexism in Current ESL Textbooks. *TESOL Quarterly, 18*(4), 705-724.

Prodromou, L. (1988). English as Cultural Action. *ELT Journal, 42*(2) 73-83.

Razmjoo, S. A. (2007). High Schools or Private Institutes Textbooks? Which Fulfill Communicative Language Teaching Principles in the Iranian Context? *The Asian EFL Journal Quarterly, 9(4),* 125-139.

Renner, C. (1997). Women are Busy, Tall, and Beautiful: Looking at Sexism in EFL Materials. From Annual Meeting of the Teachers of English to Speakers of Other Languages. Orlando Fl., March 11-15.

Reynolds, J. C. (1967). An assessment of the cognitive content of review questions in selected secondary general biology textbooks. Unpublished doctoral dissertation, University of Alabama, United States.

Riazi, A. M. (2003). What textbook evaluation schemes tell us? A study of the textbook evaluation schemes of three decades. In W. A. Renandya (Ed.), *Methodology and materials design in language teaching* (pp52-68). Singapore: SEAMEO regional center.

Sadeghi, A. R. (2005). EAP Methodology: A Transition from the Present State. Paper presented in the First National ESP/EAP Conference, Tehran, Iran.

Sadeghian, J. (2005). Developing bilingual professionals. Paper presented in the First National ESP/EAP Conference, Tehran, Iran.

Santa, C. M., & Burstyn, J. N. (1977). Complexity as an impediment to learning. *Journal of Higher Education, 48(5),* 508-518.

Schmidt, J. L. (1971). An assessment of the cognitive content of questions contained in selected high school physics textbooks. Unpublished doctoral dissertation, University of Arkansas, United states.

Shahedi, S. (2001). Constructing analytical framework for the analysis of Persian language texts for foreign learners. Unpublished master's thesis, Shiraz University, Shiraz.

Sheldon, L. (1988). Evaluating ELT textbooks and materials. *ELT Journal, 42(4),* 237-246.

Sunderland, J. (1992). Gender in the EFL Classroom. *ELT Journal. Volume, 46*(1), 81-91.

Tajeddin, Z. (2005). English for Business Communication. Paper presented in the First National ESP/EAP Conference, Tehran, Iran.

Tayebipour, F. (2005). The Role of TEFL Instructors vs. Specific-field Instructors in ESP/EAP Teaching. Paper presented in the First National ESP/EAP Conference, Tehran, Iran.

Tomlinson, B. (2001). Materials development. In R. Carter & D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp. 66- 71). Cambridge: Cambridge University Press.

Tucker, CA. (1975). Evaluating beginning textbooks. *English Teaching Forum, 13*, 335-61.

Ur, P. (1996). *A course in language teaching: Practice and theory.* Cambridge: University Press.

Vellenga, H. (2004). *Learning pragmatics from EFL & ESL textbooks.* Michigan: Michigan University center.

Williams, D. (1983). Developing criteria for textbook evaluation. *ELT Journal, 37*(2), 251-255.

Withrow, J. R. (1971). An analysis of the levels of cognitive skill required by the textbook questions in selected state histories. Unpublished doctoral dissertation, University of Virginia, United States.

Yarmohammadi, L. (2002). The evaluation of pre-university textbooks. *The Newsletter of the Iranian Academy of science, 18,* 70-87.

Yule, G., Matthis, T., & Hopkins M. (1992). On Reporting What Was Said. *ELT Journal, 46*(3), 245-51.

APPENDIX A

TEXTBOOK EVALUATION FORM

(For ESP English Textbooks)

Dear Colleague/Student

This evaluation form is prepared to collect your opinion about the ESP English textbooks currently in use. Your comments can be helpful in providing useful insights regarding teaching ESP English Courses in the coming semesters. The anonymity of the respondents is guaranteed. Please first read the syllabus of the ESP English Course and then answer the questions.

Thank you for your cooperation.

The Objective of the ESP English Course:

- To achieve the required skills in the reading and comprehension of the English texts in order to make the students prepared in learning specialized texts.

The Major Points:

- Reviewing the language structures and vocabulary items as covered in ESP textbooks,
- Reading various semi-specialized texts with reasonable vocabulary items,
- Focusing on learning skills including: the use of English dictionaries, vocabulary-building, writing and paraphrasing about the relevant topics,
- Mastering the special English structures, vocabularies and technical terms used or common in their discipline,
- Achieving good reading comprehension strategies,
- Being able to reproduce orally and discuss the specialized subjects and materials,
- Realizing various functions in the structure of the target language and role-playing in different situations frequently used in their discipline.

Criteria:

Please assign one of the numbers to the items given in the tables.

NECESSITY:
4: Highly Necessary 3: Necessary 2: Almost Necessary 1: Not Necessary

PRESENT CONDITION:
5: Excellent 4: Good 3: Adequate 2: Poor 1: Totally Lacking

1-Theoretical Considerations

Scale Items	NECESSITY (WEIGHT) 1 2 3 4	PRESENT CONDITION 1 2 3 4 5
1. The textbook is appropriate for the curriculum and fits the objectives.		
2. The textbook has clearly stated goals for proficiency in the specialized sense.		
3. The textbook provides a clear explanation of its theoretical and methodological orientation.		
4. The content of the book is consistent with the stated theoretical and methodological objectives.		

2- Organizational Features & Practical Considerations

Scale Items	NECESSITY (WEIGHT) 1 2 3 4 5					PRESENT CONDITION 1 2 3 4 5				
1. The cover of the book is appealing.										
2. The size of the book is appropriate.										
3. The book appears durable.										
4. Textbook format is visually appealing and interesting.										
5. Layout is clear and consistent.										
6. Textbook provides useful table of content, glossary and index.										
7. Appendices are accurate and easy to use.										
8. Textbook contains references, bibliography, and resources.										
9. Lessons/units are arranged logically.										
10. Lessons/units are of appropriate length and number.										
11. Diagrams, figures & pictures are clear, accurate & easy to understand.										

3- Content

Scale Items	NECESSITY (WEIGHT) 1 2 3 4 5					PRESENT CONDITION 1 2 3 4 5				
1. The course book's topical content and linguistic content (structures, patterns, and lexis) are relevant to the aims of the curriculum and syllabus.										
2. The topics and content of the textbook are appealing.										
3. Topics cover a variety of special texts.										
4. The content of the textbook is up-to-date and includes global issues.										
5. The subject matter in each unit is presented in a logical and organized manner.										
6. The linguistic content is derived from authentic texts.										
7. The content contains language specific issues that interest the reader.										

8. The content materials in the course books are designed to meet the learner's specific need.									
9. Activities are developmentally appropriate.									
10. Activities apply to a diversity of students abilities, interests and learning styles.									
11. Directions are clearly written and explained.									
12. The related technical terms and keywords are sufficiently introduced.									
13. The content is according to the student's knowledge level of their discipline.									
14. There is a correspondence between the content and student's discipline.									

4- Skills

Scale Items	NECESSITY (WEIGHT) 1 2 3 4 5					PRESENT CONDITION 1 2 3 4			
1. Activities in the textbook are balanced among various skills.									
2. An adequate number of skill-building activities are provided.									
3. Each skill is treated in every unit/ lesson.									
4. There is a balance among the various skill-building activities.									
5. The activities match the required skills that the learners need in the target situation.									
6. Skill-building activities are process-oriented (containing preview and follow-up activities).									
7. Skill-building strategies are included in the text.									
8. There is a logical/gradual development of the skills throughout the textbook.									
9. Both 'top-down' and 'bottom-up' reading strategies are suggested.									
10. Instructions in the textbook tell students to read for comprehension.									
11. The textbook contains activities which activate students' background knowledge.									
12. The textbook introduces various reading techniques (skimming, scanning, etc.).									

5- Vocabulary

Scale Items	NECESSITY (WEIGHT) 1 2 3 4				PRESENT CONDITION 1 2 3 4 5				
1. Vocabulary items are useful and frequent in the students' discipline.									
2. Vocabulary items are presented at an appropriate rate and in reasonable amounts per lesson/unit.									
3. Presentation of vocabulary is contextualized.									
4. The sequencing of vocabulary is logical and in increasing order of difficulty.									
5. The sequencing of vocabulary provides for recycling and review.									
6. Various meaningful and vocabulary practices are provided.									
7. The textbook provides appropriate techniques for learning new vocabulary items.									
8. There are interactive and task-based activities that require students to use new vocabulary.									

6- Structure

Scale Items	NECESSITY (WEIGHT) 1 2 3 4				PRESENT CONDITION 1 2 3 4 5				
1. The amount of special grammar taught in each chapter/ lesson is adequate.									
2. Grammar is presented in a logical manner and in order of difficulty.									
3. The sequencing of grammar provides for recycling / review.									
4. Grammatical structures presented in the text are accurate.									
5. Examples of grammatical structures are contextualized.									
6. Explanations of grammatical structures are clear.									

• Your **overall impression** of the textbook**: ------**

5 = excellent, 4 = good, 3 = adequate, 2 = poor, 1 = absolutely negative

Thank You

Appendix B

Students' Viewpoints Regarding ESP Textbooks

Criterion 1: Theoretical Considerations (Students)

Table 16. Descriptive Statistics for the Theoretical Considerations

Necessity	NN		AN		N		HN	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**20**	**13.6**	**34**	**23.1**	**43**	**29.3**	**50**	**34**
Present	TL		P		A		VG	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**29**	**19.7**	**52**	**35.4**	**34**	**23.1**	**32**	**21.8**

Table 17. Theoretical Considerations: Paired-sample t-test between 'necessity' and 'present' conditions

Theoretical Considerations	N	Mean	Std. Deviation	t	Sig.
Present	147	2.4694	1.04230	2.805	.006*
Necessity	147	2.8367	1.04711		

* Significant at .05

Criterion 2: Organizational Features and Practical Considerations (Students)

Table 18. Descriptive Statistics for the Organizational Features and Practical Considerations

Necessity	NN		AN		N		HN	
Condition	**F**	**P**	**F**	**P**	**F**	**P**	**F**	**P**
	17	**11.6**	**45**	**30.6**	**60**	**40.8**	**25**	**17**
Present	TL		P		A		VG	
Condition	**F**	**P**	**F**	**P**	**F**	**P**	**F**	**P**
	21	**14.3**	**31**	**21.1**	**56**	**38.1**	**39**	**26.5**

Table 19. Organizational Features and Practical Considerations: Paired-sample t-test between 'necessity' and 'present' conditions

Organizational Features	N	Mean	S.D	t	Sig.
Present	147	2.7687	1.00047	-1.233	.220
Necessity	147	2.6327	.89963		

Criterion 3: Content (Students)

Table 20. Descriptive Statistics for the Content

Necessity	NN		AN		N		HN	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**14**	**9.5**	**29**	**19.7**	**40**	**27.2**	**64**	**43.5**
Present	TL		P		A		VG	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**40**	**27.2**	**54**	**36.7**	**23**	**15.6**	**30**	**20.4**

Table 21. Content: Paired-sample t-test between 'necessity' and 'present' conditions

Content	N	Mean	S.D	t	Sig.
Present	147	2.2925	1.08040	6.669	.000*
Necessity	147	3.0476	1.00909		

* Significant at .05

Criterion 4: Skills (students)

Table 22. Descriptive Statistics for the Skills

Necessity	NN		AN		N		HN	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**14**	**9.5**	**35**	**23.8**	**39**	**26.5**	**59**	**40.1**
Present	TL		P		A		VG	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**10**	**6.80**	**20**	**13.6**	**74**	**50.3**	**43**	**29.3**

Table 23. Skills: Paired-sample t-test between 'necessity' and 'present' conditions

Skills	N	Mean	S.D	t	Sig.
Present	147	3.0204	.83968	-.426	.671
Necessity	147	2.9728	1.01324		

Criterion 5: Vocabulary (students)

Table 24. Descriptive Statistics for the Vocabulary

Necessity	NN		AN		N		HN	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**13**	**8.8**	**38**	**25.9**	**30**	**20.4**	**66**	**44.9**
Present	TL		P		A		VG	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**16**	**10.9**	**34**	**23.1**	**42**	**28.6**	**55**	**37.4**

Table 25. Vocabulary: Paired-sample t-test between 'necessity' and 'present' conditions

Vocabulary	N	Mean	S.D	t	Sig.
Present	147	2.9252	1.02093	.703	.483
Necessity	147	3.0136	1.03359		

Criterion 6: Structure (students)

Table 26. Descriptive Statistics for the Structure

Necessity	NN		AN		N		HN	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**21**	**14.3**	**20**	**13.6**	**29**	**19.7**	**77**	**52.4**
Present	TL		P		A		VG	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Condition	**32**	**21.8**	**65**	**44.2**	**22**	**15**	**28**	**19**

Table 27. Structure: Paired-sample t-test between 'necessity' and 'present' conditions

Structure	N	Mean	S.D	t	Sig.
Present	147	2.3129	1.01902	6.089	.000*
Necessity	147	3.1020	1.10871		

* Significant at .05

Overall impression (students)

28. Descriptive Statistics for the Overall impression (Students)

Principles	TL		P		A		VG	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Overall Impression	**72**	**49**	**45**	**30.6**	**23**	**15.6**	**7**	**4.8**

Comparison of Students' Viewpoints Regarding ESP textbooks

Table 29. Paired Sample t-test among students' viewpoints Concerning ESP textbooks

Student's viewpoints	N	Mean	S.D	t	Sig.
Present	147	2.6757	1.11702	4.904	.000*
Necessity	147	2.9342	1.04071		

* Significant at .05

The Asian ESP Journal

A Case Study of Technology Integration in an EFL Classroom

Zhu Ming

Jiangsu Polytechnic University

Biodata

Zhu Ming is an Associate Professor of Applied Linguistics in the School of Foreign Languages at Jiangsu Polytechnic University, China. She holds an M.A. in Applied Linguistics from Shanghai International Studies University. Her areas of research include second language acquisition, technology-enhanced language learning, TEFL pedagogy, and pragmatics. She has published the book *College English Extensive Reading* (Nanjing Normal University Press, 2006).

Abstract

This study examines the evaluation of technology integration in a Chinese English-as-a-foreign-language (EFL) classroom. It focuses on the description of the effectiveness of web-based instructional environments (WBIEs) facilitated by the WebNing course management system. Data used for analysis include an online survey and oral interviews. The study demonstrates that students perceive the use of web-based instruction (WBI) as advantageous. Findings show students feel WBI enhances their classroom experience by (1) developing writing skills through online essays; (2) improving communication skills through online discussion; and (3) providing greater learner autonomy. Furthermore, web-based learning

empowers students to be actively involved in the learning process and to be responsible for their own learning. As technology-enhanced instruction is relatively new in China, the findings of this study can provide a useful reference point for Chinese English teachers contemplating the implementation of WBI.

Keywords: technology integration; web-based instruction; Chinese EFL students

1. Introduction

Technology enhanced instruction and the integration of new technologies into the foreign language curriculum have become a growing and significant component of foreign language instruction. Online course management systems, especially asynchronous computer-mediated communication (CMC; e.g., e-mail, electronic bulletin boards) promote interactive learning, which is central to the development of communicative language skills. In addition to providing opportunities for online activities, technology has become an integral part of learning and teaching. Previous studies have documented a number of benefits to students of internet technology (e.g., Chun & Wade, 2003; Darhower, 2002; Lee, 2002, 2004; Sengupta, 2001; Smith, 2003; Warschauer, 2000). Technology-enhanced instruction offers students unparalleled access to instructional resources, far surpassing the reach of the traditional classroom. It also makes possible learning experiences that are open, flexible, and distributed, providing opportunities for engaging, interactive, and efficient instruction (Smith & Hardaker, 2000). Furthermore, cognitive theories of learning have extended the design and delivery of WBI, providing theoretical underpinnings for instructional practice (Rovai, 2002).

As web-based instruction is relatively new in China, some uncertainty exists, as little research has been conducted into the effectiveness of integrating technology in Chinese EFL instruction. Therefore this study serves as an initial examination of the empirical evidence for

the instructional effectiveness of WBI in China.

The purpose of this study is to explore Chinese EFL students' perspectives on the effectiveness of using the WebNing course management system. The study presents findings on how student roles, learning conditions, beliefs, and attitudes shape their feelings toward WBI for foreign language acquisition. First, a theoretical framework as it applies to WBI and language learning will be discussed. Second, the design of the research will be described. Finally the results of the study, based on an online survey and oral interview data, will be reported and discussed, along with suggestions for future improvement and research.

2. Theoretical Framework

CMC is a widely used educational tool because it lends itself to instruction based on sociocultural principles. It has been suggested by Vygotsky (1978) and explored by many researchers (Aljaafreh & Lantolf, 1994; Anton & DiCamilla, 1998; Coughlin & Duff, 1996; Warschauer, 2000) that learning takes place in a social environment and is facilitated by dialogue. During this exchange of ideas, each interlocutor is able to internalize new jointly constructed knowledge. Learning is therefore mediated by the context in which it takes place, the tools used to aid the learning process, such as dialogue, electronic messages, and web pages, as well as by the participants involved in the learning process. This collaborative learning process is often facilitated by scaffolding, when interlocutors provide explanations, questions, and suggestions to assist each other in solving problems (Donato, 1996). Collaborative learning tasks encourage learner autonomy, build teamwork, alter the role of teachers and students, allow students to scaffold, facilitate class discussion, and promote critical thinking (Henri & Rigault, 1996). When students provide scaffolds for each other, they can ultimately reach higher-level understanding of tasks or solve problems they would have otherwise been unable to solve alone. When done through writing instead of speaking,

the writing process changes from an independently performed task to one that promotes use of the input and reflections of other students.

Apart from the cognitive benefits described above, students and teachers alike have witnessed the positive social impact CMC can have. Asynchronous electronic exchanges seem to foster the building of a learning community, where participants offer each other support and praise (Cole, Raffier, Rogan, & Schleicher, 1998; McKenzie & Murphy, 2000; Sengupta, 2001). Findings illustrate how "CMC creates the opportunity for a group of people to construct knowledge together, thus linking reflection and interaction" (Warschauer, 1997, p. 473). Therefore, it is a good fit for promoting the type of student-centered learning central to the sociocultural theory of language learning.

Within this theoretical framework, the project described below advances the research that examines technologically-enhanced instruction. The multichannel web tools of WebNing allow teachers to create online activities which promote interaction and collaboration among students.

3. The Study

This study investigates whether participation in a one-semester network-based course affects student attitudes and beliefs toward the development of their English skills and the use of Internet technology.

3.1 Participants

In the second semester of the academic year 2007, a university English class in South China consisting of sixty-one freshman undergraduate students majoring in computer science,

participated in the project. They had completed six years of formal English education before studying at the university. At the outset, all the participants were required to use WebNing to support their English learning. In terms of computer access and literacy, students can do their course work in the computer learning center or on their own computers and had used New Horizon College English online prior to the study. Therefore, most of the students were familiar with how to operate the online system.

3.2 Description of the Project

The project was part of the required curriculum for College English and it was developed to help students improve their English competency with the aid of internet technology. Primary teaching methods involved face-to-face instruction (four hours per week) and a web-based component which consisted of the following sections: course information (syllabus, lecture notes), assignments, student page, group page, discussion forum, and blogs. For this project WebNing provided two major instructional tools. One was the content tool which delivered course materials, and the other was the communication tool which allowed users to interact with each other in cyberspace using a forum and blogs (for more information about WebNing, visit http://onlineenglish2008.ning.com, password: zg711=15). The pedagogical objectives of the WebNing-based course were (1) to ensure all the students interact with course materials outside lecture time and (2) to increase student-teacher and student-student interaction in the web-based instructional environments.

In the first week of class, the instructor informed students that they were required to use WebNing to carry out a variety of online tasks in order to develop their English skills throughout the semester. Within the first two weeks of the course, the students were asked to use their e-mail account and password to login to access course materials, and to use the forum and blogs as part of their class participation credit. The online activities were scheduled

in the course syllabus and available online. In order to avoid the possibility of confusion and to ensure students knew how to complete online assignments, the instructor demonstrated the major tools they would be using.

3.3 Online Essay

Every week students were required to write an essay to answer open-ended questions. Questions were based on topics related to the course textbook or daily life, such as environmental awareness or advantages and disadvantages of studying abroad. Posted online were a set of different topics, each with a dozen open-ended questions for essay writing (sample questions are included in Appendix A). The students posted their completed assignments in blogs. All the students could comment on the essays in the blogs. The instructor then corrected the essays by writing comments and correcting mistakes on usage, vocabulary, and grammar. For instance, the instructor underlined errors and wrote the code for the correction, such as "vt" for verb tense, "voc" for wrong word, and "prep" for incorrect preposition (Ailing & Xiangming, 1997). The assignments which had been reviewed by the instructor were then returned to the students through the blogs. The students retrieved the documents and then responded to the instructor's feedback and corrected the errors. Finally, they downloaded the revised copy and placed it in their page as part of the course requirement.

3.4 Online Discussion

In addition to online essay writing, students were also required to post their opinions and responses to the discussion topics, and discussion assignments were posted every two weeks. The instructor used an asynchronous forum board to engage students in an online discussion activity outside of class time. The aim of online discussions was to allow students to freely

express themselves about their opinions and to exchange ideas through informal social interaction on the assigned topics (see sample online discussion questions in Appendix B). Discussions were based on such topics as the role of men and women in modern society or young adulthood. The instructor divided the class into small groups for the purpose of enhancing the discussions. Several threads were set up in the forum board. Students were aware that there was a regular moderating presence, and the instructor frequently accessed the discussion forums, posted comments, and answered questions. To enhance discussion, instructions and rules were developed by the instructor and posted to the students (see Figure 1). According to a study by Beaudin (1999), providing guidelines for an online activity should be rated as an important criterion for keeping the discussion on-topic. Discussion forums can replace the casual conversations that take place between students and they have the benefit of creating a permanent written log. Clark (2001) points out it is important that "students have ample time to read other students' comments, do research, and formulate a detailed response" (pp. 119-124).

Table 1. Rules for Online Discussion

1	Do not be afraid of posting the wrong answer—this is a discussion, not a test.
2	All messages must be relevant to the topic discussion.
3	All the students must participate—i.e., lurking is not allowed.
4	Before posting make sure that your messages are concise and grammatically correct.
5	All dialogue is archived for you to return to for reflection at a later time.
6	Each group should nominate a group leader and only that member of the group should be logged in.

4. Results and Discussion

4.1 Online Survey Results

In order to understand general student attitudes toward the WBI, within two weeks of completion of the projects described above, all students were required to complete a questionnaire and short-answer questions posted on WebNing.

The questionnaire contained 18 statements which students were asked to indicate whether they strongly agreed, agreed, disagreed, or strongly disagreed. These statements elicited information about students' views in four categories: 1) perspectives on web-based instructional environments, 2) the effect of WBI on learning subject matter and skills, 3) the effect of WBI on their interest, and 4) the effect of WBI on their confidence as learners, their technical skills, and their performance on course assessments.

The questionnaire data were reported in percentages of student responses for each statement, as shown in Appendix C. The statements were ordered from those receiving the highest number to the lowest number of responses expressing agreement.

4.1.1 Perspectives on Web-Based Instructional Environments

WebNing created an active and supportive instructional environment in which students interacted socially with each other as they expanded their use of English beyond the classroom. Most of the students (81.3%) praised the unique learning community created by WebNing, and said they enjoyed using WBIEs. Over 60% reported they had more interaction and communication with the instructor, as well as with their classmates. They acknowledged WebNing made communication efficient and effective because they were able to receive responses from their classmates and constructive feedback from the instructor without delay.

More than half (58%) felt free to ask questions to the instructor. Being able to work and share opinions with others was a valuable added benefit of online exchanges. When asked about ease of use, 68% (combined response percentage of "agree" and "strongly agree") of the students reported they were comfortable using it.

4.1.2 Perspectives Concerning the Effect on Learning

After a semester of using WBI, 70% of the students felt they learned more language skills than they would have learned in a regular class. 60% stated their writing skills improved because of the online writing activities. In addition, a slight majority of students (52%) found the writing topic-based online essays enjoyable.

The asynchronous discussion was the most innovative aspect presented in this project. Regarding the use of the communications tools, student comments were overwhelmingly positive. This positive evaluation was also substantiated in comments during participation in a voluntary online discussion (see Table 2).

Table 2. Question: What Do You Think of Online Discussion?

Positive Comments	Negative Comments
We would not have discussions like this in the regular class.	I find it difficult to get my message across without speaking face-to-face.
Interesting and good fun.	Big group can be confusing.
Other students' views are important and they give you more ideas and can broaden your vision.	Time consuming. I need face-to-face for communication.
Good and remarkably fast discussion.	It will be better with a small group.

Fun and a good learning experience. Let's have more of these activities.	
You certainly learn from it.	

The questionnaire showed that half the students (50%) felt they learned a considerable amount from the online discussion activity. Linguistically, students felt that the process of composing ideas allowed them to recycle vocabulary and structures they had previously learned as they read and responded to their classmates. Cognitively, they felt that they were challenged by their limited linguistic skills and that they had to stop and think about what they would say and need to say in order to get their ideas across to their peers.

Finally, only 28% of the students believed working with their classmates contributed significantly to the development of their communicative skills. Conrade (1999) finds that when compared to fourth semester students, first semester language students place a lower priority on developing communicative skills. Since the students in this project were first year students, this may explain this finding.

4.1.3 Enjoyment and Interest

According to questionnaire results, the majority of students (73%) agreed WebNing made the course more interesting, and two-thirds (66%) said that if given a choice between a regular English class and a WBI, they would take the latter. In addition, more than half (53%) reported that they enjoyed doing the online assignments more than traditional assignments.

Interestingly, these findings coincide with the online written feedback where the majority of students spoke favorably about the WBI (see Figure 3), describing it with 52 positive comments, compared to 3 negative ones. Among positive comments, the most

common were "interesting / interested" and "good". Students attributed 56 strengths to the project, while listing only 16 weaknesses.

Figure 3. Question: What Do You Think of This Project?

A. Adjectives	**Number**	**C. Likes**	**Number**
Interesting / Interested	15	Using computers	10
Good / Nice	10	Group discussion	9
Boring	3	E-mail	9
Enjoyable / liked	5	Essay writing	8
Exciting	4	Cooperative learning	4
Fresh / new	4		
Exciting / wonderful / great	6	**D. Dislikes**	
Useful	2	Didn't like group discussion	4
Encouraging	3	Didn't like learning by computer	4
meaningful	3	Didn't like essay writing	3
B. Strengths		**E. Weaknesses**	
Improved English	16	Not enough test-related skills	5

Improved writing	11	Too time consuming	4
Improved grammar	6	A lot of work	3
Improved communication	6	Students lack computer skills	3
A good learning experience	5	Many new vocabulary	1
Improved confidence	3		
Improved computer skills	3		
Improved typing skills	3		
Learned from others	3		

In the questionnaire, when asked if web-based learning could take the place of live lectures, 66% stated that they found the online learning experience worthwhile but not as a replacement or alternative to face-to-face contact, but rather as a supplementary learning tool. More positive results may have emerged if the word "alternative" had been omitted, possibly replacing it with "additional". This emphasizes the need to clarify the rationale for using WBI with students, ensuring they understand that the aim is to supplement and enhance their learning and not to replace face-to-face interaction.

4.1.4 Perspectives Concerning Effects on Confidence of Learners, Technical Skills, and Class Assessments

For the majority of students, WBI appeared to have had certain benefits. 79% of the students reported that they gained confidence in their ability to complete the online activities. A majority (54%) also believed they gained confidence in their ability to use technology. These

findings corroborate research showing that WBI helps students gain confidence in language use as well as in the use of computers (Beauvois, 1994; Cononelos & Oliva, 1993; Lunde, 1990).

In addition, almost two-thirds (65%) expressed increased confidence as independent learners. This finding supports results from Blackstock Junior High (1993) that concludes that students engaged in technological interactive learning environments become independent "knowledge navigators".

As for class assessment, only 41% of the students stated their WBI experiences helped them improve their scores. The less enthusiastic perspectives concerning the effect of WBI on their scores may be due to the fact that there was no direct link between the individual online components and classroom assessments and structure. In China, students tend to confine their work to what is specifically taught in the course and they have tremendous pressure to pass a competitive examination system. In general, university students are required to take a unified general English examination at the end of the semester. Before graduating from a university, they are required to pass the National College English Test Band 4 (CET-4), which is an authoritative English proficiency test for non-English majors. The test is focused on grammar, vocabulary, listening, reading, translation, and writing. Since formative assessment tasks for the unit of study were not included in WebNing, nearly 60% of students did not believe the WBI experience could help them pass course examinations and CET-4. These findings coincide with Nowaczyk's (1998) and Warschauer's (1996) research. Nowaczyk (1998) finds that students attribute learning value primarily to those multimedia components that are directly related to course examinations, regardless of the intrinsic learning value of the material. Warschauer (1996) also finds that the degree to which computer-based projects are

integrated into general course goals and structure correlates with differences in student motivation.

4.2 Oral Interview Results

In addition to the online survey, an oral interview was given by the course instructor to students at the end of semester. During the interview, students were asked to describe their experiences with the online activities and to explain the most valuable, interesting, and difficult aspects of WBI. All of the interviews were conducted in English and recorded for later analysis. Analysis of the interviews revealed that the majority of students regarded WBI as beneficial to their language development. However, they made some suggestions for future improvements. The following discussion highlights important issues that students expressed regarding WBI for English learning.

4.2.1 Development of Writing Skills through Online Essay Writing

The goal of online essay writing was to build proficiency through describing, narrating, explaining situations, and expressing opinions in different contexts. After one semester of writing practice, the majority of students felt that their writing skills had improved. The following comments made during the interview revealed students' rewarding experiences with the online essay writing:

> Now I am more confident with my writing through writing so much in WebNing.

> As for writing, I don’t feel like I have a big problem, and I think I have lots of ideas on everything, which helps me a lot to write. But then looking at other people's ideas, I think

> mine is too narrow. Other people's perspective broadens my mind, and I think I really see things more in-depth and widely.

> At the beginning, I have some problems with the grammar and vocabulary. I spend a lot of time looking for words in the dictionary. I feel that I write better now and I'm able to express my opinions easily.

Students agreed that at the beginning of the semester weekly writing involved a lot of work. By the end of the semester, many of them were proud of how much and how easily they could write. During the interview, one student recalled her change of attitude towards the writing requirement, commenting:

> At the beginning I am skeptical about using WebNing for writing assignments. But after a few weeks, I get used to it and find the online writing very rewarding. The topics are very interesting to me. I have never written so many essays in English. I believe my writing skills have improved.

Regarding the topics for online essays, students found open-ended questions stimulating and meaningful because they were able to express their voices through their own life experiences. A few students, however, experienced difficulty in being able to fully communicate in the target language. One linguistically weak student admitted during the interview that:

> The topics are too difficult to me. I feel frustrated not being able to express myself in English. The writing process is overwhelming for me. It takes too much of my time

> writing essays. I often end up translating my ideas from Chinese to English. I think I need a lot of help.

This student further suggested that the instructor should provide students with more guidance on how to write English essays.

4.2.2 Enhancement of Communication Skills through Online Discussion

According to Lamy and Goodfellow (1999), a task that simply requires social conversation may not stretch learners' ability with language production. Pica, Kanagy, & Falodun (1993) emphasize that tasks without a specified goal may not push learners to use their linguistic resources for negotiation of meaning. The goal of online discussion is to engage students in sharing, exchanging, and debating information relevant to their life experiences. Participating in this form of social and text-based interaction is a modern enactment of Vygotsky's (1978) idea of learning as a social-cultural process where language is an essential vehicle for development. WBIEs not only encourage students to use English but also reinforce their thinking skills and communication strategies, which are crucial for successful online interaction. During the interviews, students praised the opportunity to use the target language in a natural way and they felt that their communication skills in English improved after participating in the online discussion. Students made the following comments:

> I like online discussion because we can share our ideas. It is necessary to think in English more rapidly than when you write the homework. This is more like the real world.

> Through online communication, I realize that I need to enlarge my English vocabulary because on many occasions I do not know how to say certain words and I have to look them up in the dictionary. I find it very interesting to learn English by using English.

Online discussion is perhaps a less intimidating environment than the classroom, and it may help encourage shy students to express their ideas without feelings of embarrassment. There are no anonymous postings to ensure that every student participates, but nevertheless these alternative methods provide a range of opportunities for students to participate who might not necessarily contribute in face-to-face interaction. As it is important to provide a range of learning opportunities to try to engage all students, online discussions perhaps provide an approach to engage those students who prefer not to ask questions in the classroom in front of their classmates and the instructor. One student during the interview expressed her feeling of ease communicating online:

> I hate speaking in front of people, but I feel comfortable chatting with my classmates online. Although I am very slow in composing my ideas, I do not feel nervous in the virtual learning environment. Now I'm more comfortable expressing myself in English. WebNing is very helpful and valuable for me. It motivates me to speak more in class.

WBIs offer affective support to shy students so they can carry out shared tasks without feeling pressure as intensely as they do in front of the class. In other words, WBIs build student confidence and encourage those who are reluctant to participate in oral discussions to speak (Chapelle, Compton, Kon, & Sauro, 2004; Lee, 2002). Furthermore, Sweeney and Ingram (2001) demonstrates that Asian students with weaker English language skills have a more positive evaluation of bulletin board discussions as they offer more space for them to prepare their opinions and responses.

4.2.3 Improvement of Learner Autonomy

All of the participating students felt that the experience was interesting and rewarding. Although they had prior experience using computers for learning and communication, it was the first time they used computers for discussing and writing in English and they found it exciting. As one of the students exclaimed, "English suddenly becomes familiar and close to me. This is so different from high school English classes!" One reason they like it is because WebNing organized and presented course materials logically through a single shared space, which facilitated learning. Most students quickly adapted to it in completing and checking their assignments online. More importantly, they learned how to organize their work by retrieving documents from the blogs and placing them into their page. The following comments from the interview demonstrate how students felt WebNing enhanced their organizational skills and motivated them to become independent and self-directed learners:

> Frankly speaking, I'm not a very organized person. After sending and receiving so many essay assignments from the "Blogs", I couldn't find them to make revisions. In order not to get lost, I learn to place everything in the order. WebNing makes me organize things better.

> I think it's easy to get lost if you don't know where things are. As we have to do so many things through WebNing, I have to learn how to organize them in a logical manner.

A few students, however, remarked that it was confusing and easy to get lost at the beginning because of the abundance of information presented in WebNing. Each document that students sent to the instructor's blog showed the time and date. Several students had difficulty keeping up with their work and complained about the amount of work required for

the course, often completing their homework at the last minute in the online discussions. One student admitted that he learned to organize his work by making plans and not falling behind in the course schedule. This student further explained that WebNing forced him to become efficient and self-disciplined. These comments imply learner self-sufficiency and independence have played a crucial role in web-based instruction. In other words, students need to learn to be in charge of their own learning and they need to become actively involved in the learning process, an observation which previous studies support (Lamy & Goodfellow, 1999; Lee, 2002).

5. Conclusion

This study presents the experiences and perspectives of one group of Chinese EFL students in regard to the integration of technology into English language instruction. While a one-semester investigation is limited in scope and depth, the project has succeeded in making many students understand what is required to be successful when using WBI. The use of WebNing as an instructional tool has apparently supported an active learning environment, contributing to the development of student language skills and reinforcing cognitive skills. Students stated that they benefited from the online discussions and writing practice as they were motivated by the structured online tasks and they expanded their use of the target language beyond the classroom setting. This study also reveals that online learning empowered students to be actively involved and responsible for their learning. Furthermore, WBI facilitated communicative interaction between students and teacher.

There are several limitations to this study, including the fact that student information is self-reported and self-assessed. In addition, the duration of the study was relatively short. Finally, this study focused on a small number of participants and therefore data may not

correspond to other foreign language classrooms. Nevertheless, since little research is available on Chinese student perspectives on English language learning in WBI, this study may provide insights for universities currently implementing this system.

More empirical studies should be conducted to examine student perspectives on synchronous interaction via discussion boards in WBIs. In addition, studies should be conducted concerning the role of the teacher in WBIs in order to identify teacher behaviors and interactions favorable to language acquisition. Finally, studies that include online exchanges with western peers in WBIs could advance knowledge of how students perceive the role of interaction in a socially collaborative context. Such studies may contribute to a future knowledge base that will shape and improve curriculum and instruction mediated by technology.

References

Ailing, Q. & Xiangming. Z. (1997). *Writing Skills for English Majors*. Qingdao: Qingdao Press.

Aljaafreh, A., & Lantolf, J. (1994). Negative feedback as regulations and second language learning in the zone of proximal development. *The Modern Language Journal, 78*(4), 465-483.

Anton, M., & DiCamilla, F. (1998). Socio-cognitive functions of L1 collaborative interaction in the L2 classroom. *The Canadian Modern Language Review*, *54(*3), 314-342.

Beaudin, B.P. (1999). Keeping online asynchronous discussion on topic. *Journal of Asynchronous Learning Networks, 3*, 41-53.

Beauvois, M. (1994). E-talk: Attitudes and motivation in computer-assisted classroom discussion. *Computers and the Humanities, 28*(1), 177-190.

Blackstock Junior High School. (1993). Multimedia technology drives smart school. *Technology and Learning*, *14*(1), 41-44.

Chapelle, C., Compton, L., Kon, E., & Sauro, S. (2004). Theory, research & practice in CALL: Making the links. In L. Lomicka & J. Cooke-Plagwitz (Eds.), *Teaching with technology* (pp.189-208). Boston: Thomson Heinle.

Chun, D., & Wade, E. R. (2003). Collaborative cultural exchanges with asynchronous CMC. In L. Lomicka & J. Cooke-Plagwitz (Eds.), *Teaching with technology* (pp. 220-247). Boston: Thomson Heinle.

Clark, J. (2001). Stimulating collaboration and discussion in online learning environments. *Internet and Higher Education, 4*, 119-124.

Cole, R., Raffier, L.M., Rogan, P., & Schleicher, L. (1998). Interactive group journals: Learning as a dialogue among learners. *TESOL Quarterly*, *32*(3), 556-568.

Cononelos, T., & Oliva, M. (1993). Using computer networks to enhance foreign language/culture education. *Foreign Language Annals, 26*(4), 494-512.

Conrade, D. (1999). The student view on effective practices in the college elementary and intermediate foreign language classroom. *Foreign Language Annals*, *32*(4), 494-512.

Coughlin, P. & Duff, P. (1996). Same task, different activities: Analysis of SLA task from an activity theory perspective. In Lantolf, J. & Appel, G. (Eds.), *Vygotskian Approaches to Second Language Research* (pp. 173-193). New Jersey: Ablex Publishing Corporation.

Darhower, M. (2002). Interactional features of synchronous computer-mediated communication in the intermediate L2 class: A sociocultural case study. *CALICO Journal*, *19*, 249-277.

Donato, R. (1996). Collective scaffolding in second language learning. In J. Lantolf & G. Appel (Eds.), *Vygotskian Approaches to Second Language Research* (pp. 33-56). New Jersey: Ablex Publishing Corporation.

Henri, F. & Rigault, C.R. (1996). Collaborative distance learning and computer conferencing. In T. Liao (Ed.), *Advanced Educational Technology: Research Issues and Future Potential* (pp. 45-76). Berlin: Springer Verlag.

Lamy, M.N., & Goodfellow, R. (1999). Reflective conversation in the virtual language classroom. *Language Learning & Technology*, *2*, 43-61. Retrieved May 1, 2005, from http://llt.msu.edu/vol2num2/article2

Lee, L. (2002). Enhancing learners' communication skills through synchronous electronic interaction and task-based instruction. *Foreign Language Annuals*, *35*, 16-23.

Lee, L. (2004). Learners' perspectives on networked collaborative interaction with native speakers of Spanish in the US. *Language Learning & Technology*, *8*, 83-100. Retrieved May 1, 2005, from http://llt.msu.edu/vol8num1/lee/default.html

Lunde, K. (1990). Using electronic mail as a medium for foreign language study and instruction. *CALICO Journal*, *7*(3), 68-78.

McKenzie, W., & Murphy, D. (2000). "I hope this goes somewhere": Evaluation of an online discussion group. *Australian Journal of Educational Technology*, *16*(3), 239-257. Retrieved March 5, 2004, from http://www.ascilite.org.au/ajet/ajet16/mckenzie.html

Nowaczyk, R. (1998). Student perception of multimedia in the undergraduate classroom. *International Journal of Instructional Media*, *25*, 367-368.

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

Rovai, A.P. (2002). Sense of community, perceived cognitive learning, and persistence in asynchronous learning networks. *Internet and Higher Education*, *5*, 319–332.

Sengupta, S. (2001). Exchange ideas with peers in network-based classrooms: An aid or a pain? *Language Learning & Technology*, *5*, 103-134. Retrieved May 1, 2005, from http://llt.msu.edu/vol5num1/sengupta/default.html

Smith, B. (2003). Computer-mediated negotiated interaction: An expanded model. *Modern Language Journal*, *86*, 3-57.

Smith, D., & Hardaker, G. (2000). e-Learning innovation through the implementation of an Internet supported learning environment. *Educational Technology and Society*, *3*, 1–16.

Sweeney, J.C., & Ingram, D. (2001). A comparison of traditional and web-based tutorials in marketing education: An exploratory study. *Journal of Marketing Education*, *23*(1), 55-62.

Vygotsky, L. (1978). *Mind in Society.* Cambridge, MA: Harvard University Press.

Warschauer, M. (1996). Motivational aspects of using computers for writing and communication. In M. Warschauer (Ed.), *Telecollaboration in foreign language learning: Proceedings of the Hawai'i symposium* (Technical Report #12; pp. 29-46). Honolulu: University of Hawai'i, Second Language Teaching & Curriculum Center.

Warschauer, M. (1997). Computer-mediated collaborative learning: Theory and practice. *Modern Language Journal, 81*(4), 470-481.

Warschauer, M. (2000). On-line learning in second language classroom: An ethnographic study. In M. Warschauer & R. Kern (Eds.), *Network-based language teaching: Concepts and practice* (pp. 41-58). Cambridge, UK: Cambridge University Press.

Acknowledgements

This research was supported by a Grants-in-Aids for Humanities and Social Sciences Research from the Educational Department of Jiangsu Province (No. 08SJD7400005). Many thanks are extended to the anonymous Asian ESP Journal reviewers whose insightful comments helped strengthen this article.

APPENDIX A: Sample Questions for Online Essays

Week 2:

After finishing Unit 2, you may have more knowledge of the environmental problems. 1) Could you tell us which issue concerns you and rank them in order of importance: a) river and sea pollution; b)Antarctica; c) destruction of the ozone layer; d) whaling; e) pollution from cars; f) nuclear power; g) nuclear waste; h) acid rain; i) the greenhouse effect.

2) Are you satisfied with the government's performance on environmental matters?

3) Have you noticed deterioration in your local environment in the last ten years?

4) Please tell us about other environment anxieties which may be concerning you.

Week 4:

If you are given a chance to study abroad, where do you want to study? Why? What do you want to study? Why? What might be the advantages and disadvantages of studying abroad? Do you think our government should encourage Chinese to study abroad?

APPENDIX B: Sample Questions for Online Discussion

Topic 1: An ideal youth

What is an ideal youth for you? Exchange ideas with your classmates. Do you think young people nowadays have too much responsibility and little freedom? Explain your reasons.

Topic 2: The role of women and men in today's society

What are the roles of women and men in today's society? Do women have the same rights as men? Explain your reasons.

APPENDIX C: Summary of the Questionnaire

Category 1: Perspectives on Web-based Instructional Environments (WBIEs)

	Strongly	Disagree	Agree	Strongly	No
	%	%	%	%	%
5. I think WBIEs were useful to English learning	5.6	13.1	52.8	28.5	0.0
11. I felt comfortable in the WBIEs	7.3	22.9	49.2	18.4	2.2

	Strongly	Disagree	Agree	Strongly	No
	%	%	%	%	%
4. I learned more English language skills than I would have learned in a regular English class	6.4	20.7	33.2	36.3	1.4

18. I had more interaction and communication with the instructor in the WBIEs	16.8	19.6	45	18.4	0.3
9. I had more interaction and communication with my classmates in the WBIEs	14.0	25.4	48.3	11.7	0.6
14. I felt free to ask questions in the WBIEs	14.8	26.4	43	15.3	0.6

Category 2: Perspectives Concerning the Effect on Learning

12. I believe online essays have improved my writing skills.	13.4	25.8	40.2	19.8	0.8
15. I enjoyed writing topic-based online essays	16.2	31.3	39.9	12.3	0.3
17. I think online discussion has enhanced my communicative skills	13.4	35.8	40.2	9.8	0.8
8. Working with someone online was beneficial to me	26.8	42.5	21.8	6.1	2.8

Category 3: Perspectives on Interest and Relevance

	Strongly	Disagree	Agree	Strongly	No
	%	%	%	%	%
1. The web-based learning experiences made this a more interesting course	8.7	17.6	52	21.2	0.0
3. If given a choice between a regular English class and a WBI, I would take the latter	10.6	23.5	50.3	15.6	0.6
10. I enjoyed doing the online assignments better than traditional assignment	20.1	24.3	39.7	13.7	2.2
7. The tasks I performed on the web were interesting	22.1	27.1	43.9	6.7	0.3
6. I found learning in a web-based instructional environment a worthwhile experience as an alternative to face-to-face lecture	22.1	43.9	28.8	4.2	1.1

	Strongly	Disagree	Agree	Strongly	No
	%	%	%	%	%
2. I was confident about completing assignments through the web	4.5	16.5	63.4	15.6	0.0
13. I have become a better problem-solver after using the computer while learning English	6.1	28.2	52.8	12	0.8
16. I gained confidence in my ability to use technology successfully	14.8	29.6	44.1	10.3	1.1
6. I was confident to get a good grade in the course	22.1	34.4	32.7	8.1	2.8

Category 4: Perspectives Concerning Effect on Confidence as a Learner, Technical Skills and Class Assessments